

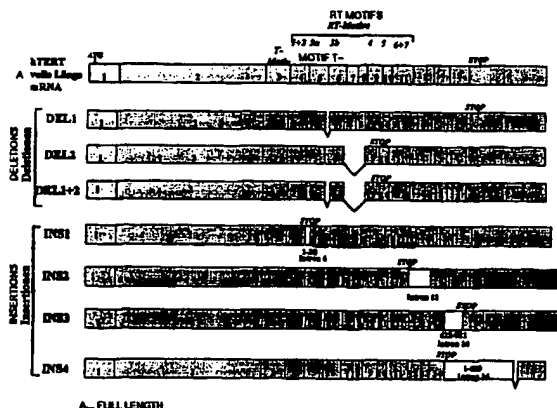
**PCT**  
 WELTORGANISATION FÜR GEISTIGES EIGENTUM  
 Internationales Büro  
 INTERNATIONALE ANMELDUNG-VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
 INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



<p>(51) Internationale Patentklassifikation <sup>6</sup> :  <b>C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027</b></p>	<b>A2</b>	<p>(11) Internationale Veröffentlichungsnummer: <b>WO 99/33998</b></p> <p>(43) Internationales Veröffentlichungsdatum: <b>8. Juli 1999 (08.07.99)</b></p>
<p>(21) Internationales Aktenzeichen: <b>PCT/EP98/08216</b></p> <p>(22) Internationales Anmeldedatum: <b>22. Dezember 1998 (22.12.98)</b></p> <p>(30) Prioritätsdaten:  <b>197 57 984.1      24. Dezember 1997 (24.12.97)    DE</b></p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): <b>BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).</b></p> <p>(72) Erfinder; und          (75) Erfinder/Anmelder (nur für US): <b>HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen-dorfstrasse 59, D-51061 Köln (DE).</b></p> <p>(74) Gemeinsamer Vertreter: <b>BAYER AKTIENGE-SELLSCHAFT; D-51368 Leverkusen (DE).</b></p>	<p>(81) Bestimmungsstaaten: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Veröffentlicht</b>  <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>	

(54) Title: **REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF**

(54) Bezeichnung: **REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG**



(57) Abstract

The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

**LEDIGLICH ZUR INFORMATION**

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen  
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-  
dung

5     Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten  
10    Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15    Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20    Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25    Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der  
30    Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

15

Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

#### Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

30

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

#### 10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34<sup>+</sup>38<sup>+</sup>-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischem Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5       Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase  
10       enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.  
15

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweiligen Strahlen schädigen nicht nur die Tumorzellen,  
20       sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle  
25       Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden  
30       dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

#### Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-translationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

30

5  
schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

10  
Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

15  
Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

20  
Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

25  
Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

30  
Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).



Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon  
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere  
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).  
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- $\alpha$ , IFN- $\beta$ , IFN- $\gamma$ , TNF, TNF- $\alpha$ , TNF- $\beta$ , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls  
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

5 RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 $\alpha$ , - $\beta$ ), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

10 Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

15 Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle  $\beta$ -Glukuronidase, pflanzliche  $\beta$ -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle  $\beta$ -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

20 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

25 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

30 Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC),  $\beta$ -Galaktosidase ( $\beta$ -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH),  $\beta$ -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch  
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch  
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen,  
15 zen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.  
20

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.  
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

30 Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF  $\beta$ , Insulin oder Insulin-like Growth Factor (IGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

30

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen  
5 Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen,  
15 die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der  
20 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

30

- 5 A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;
- 10 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

15 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

25 A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

30 B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten  $\lambda$ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene  $\lambda$ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des  
10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei  $\lambda$ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten  $\lambda$ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

15 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene  $\lambda$ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur  
20 14: nicht verdaut.

B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

30

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

15

20

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

25

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

30

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymststellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch



die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

30

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

15

20

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

30

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in

10 Intron 2.

Tab. 1

3' Acceptor Sequence				5' Donor Sequence			
Intron	Exon	Exon No.	bp	Intron	Exon	Intron	bp
						on	
						No.	
5' flanking Region							
caggcgcttcccccgag	GTTCAGGCAGCGCTGCGT	1	281	CGCCCTCTCTTCCGCCAG	gtggcctccccgggtcg	1	104
catgtccttctcgttttag	GTGTCCTGCTGAAGGAGC	2	1354	TGGCTGCGCAGGAGCCAG	gtgaggaggtggtggcgt	2	8616
gaggggctctctattttag	GGTTGGCTGTGTTCCGGC	3	196	TGCAAGCATTGGAAATCAG	gtactgtatccccacgcca	3	2089
cccatgtgtcccccgag	ACAGCACTTGAAGAGGGTG	4	181	GTTCCGCAGAGAAAAAGAGG	gtggtgtgtctttgttta	4	687
ctcgcctccactcacadag	GCCGAGGCTCTCACCCTGA	5	180	TGAGCTGTACTTTGTCAAG	gtgggtgccccgggaacccc	5	494
ccctctcccttgccgggtag	GTGGATGTGACGGGCGCGT	6	156	CAAGGCTTCAAGAGCCAC	gtgaagttcacgtgtgata	6	>4660
ctcccgctgtgttttcgtag	GTCTTACCTTGACAGACC	7	96	TGCCGTGCTCATCGAGCAG	gtctgggcactgcccctgca	7	980
ctgtgtcttccccgccdag	AGTCCTCCTGTAATGAGG	8	86	CCGTGCGCATCAGGGGCAA	gtgagtcagggtggccaggt	8	2484
gtattttcccttatttttag	GTCTACGTCCAGTGCCAG	9	114	CGGGATTCGGCGGGACGG	gtgaggcctctcttcccc	9	1984
cattgccccctctgccttag	GCTGCTCCTGCGTTTGGTG	10	72	ACGCGAAAACTTCTCTCAG	gtgagggccgtgcccgtgtg	10	1871
attccccctctgtctctdag	GACCTGCTCCGAGGTGTC	11	189	TGCAGAGCGACTACTCCAG	gtgagcgcaactggccgga	11	3804
tctttatttgggactcttag	CTATGCCCGGACCTCCATC	12	127	CCTGTTTCTGATTTGCAG	gtgagcaggctgatgtca	12	880
ctgtcgcgcaatctctctdag	GTGAACAGCCTCCAGACGG	13	62	TCCTGCTGCAGCGGTACAG	gtgagccgcccaccaaagggg	13	3187
agcctctgttttcccccdag	GTTTCACGCATGTGTGCTG	14	125	CTGAAAGCCAAAGAACGAG	gtatgtcagggtgctctggc	14	781
tctgtattttggcccccgtag	GGATGCTGCTGGGGGCCAA	15	138	CTGGGCTCACTCAGGACAG	gcaagtgtgggtggaggcc	15	536
	CCCAGACGCAGCTGAGTCG	16	664	TTTTTCAGTTTGTGAAAAA	3' flanking Region		

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

### Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

10

#### Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

15

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

20

25

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

30

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5     Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca  $1,5 \times 10^6$  Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa  
10     500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml  
15     Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit  $1,5 \times 10^6$  cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte  
20     durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*  
25     (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen  
30     Experimenten jeweils 1 bis  $1,5 \times 10^6$  Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach  $\lambda$ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklont (Beispiel 4).

### Beispiel 3

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde  $\lambda$ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte  $\lambda$ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

### Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.



Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.

Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10 Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

15 20 25 30 kloniert.

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten  
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking  
10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

15 Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten  
20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch  
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

## Contig1:

	ACTTGAGCCC	AAGAGTTC	GGCTACGGT	AGCCATGATT	GCAACACCAC	ACGCCAGCCT	TGGTGACAGA	70
	ATGAGACCCT	GTCTCAAAAA	AAAAA	AATTGAAATA	ATATAAGCA	TCTTCTCTGG	CCACAGTGG	140
5	ACAAACCAG	AAATCAACAA	CAAGAGGAAT	TTTGAAACT	ATACAAACAC	ATGAAATTA	AACAATATAC	210
	TTCTGAATGA	CCAGTGAGTC	AATGAAGAAA	TTAAAAAGGA	AATTGAAAAA	TTTATTTAAG	CAATATGATA	280
	CGGAAACATA	ACCTCTCAAA	ACCCACGGTA	TACAGCAAAA	GCAGTGCTAA	GAAGGAAGTT	TATAGCTATA	350
	AGCAGCTACA	TCAAAAAAGT	AGAAAAAGCCA	GGCGCAGTGG	CTCATGCGCTG	TAATCCCAGC	ACTTTGGGAG	420
	GCCAAGGCGG	GCAGATCGCC	TGAGGTCAAG	AGTTTCAGAC	CAGCCTGACC	AACACAGAGA	AACCTGTGCG	490
10	CTACTAAAA	TACAAATTA	GCTGGGCATG	GTGGCACATG	CCTGTAATCC	CAGCTACTCG	GGAGGCTGAG	560
	GCAGGATAAC	CGCTTGAACC	CAGGAGGTGG	AGGTTGCGGT	GAGCCGGGAT	TGCGCCATTG	GACTCCAGCC	630
	TGGGTAAACA	GAGTGAAACC	CTGTCTCAAG	AAAAA	AAGTAGAAAA	ACTTAAAAAT	ACAACCTAAT	700
	GATGCACCTT	AAAGAACTAG	AAAAGCAAGA	GCAAACTAAA	CCTAAAAATTG	GTAAGAGAAA	AGAAATAATA	770
	AAGATCAGAG	CAGAAATAAA	TGAAACTGAA	AGATAACAAT	ACAAAAGATC	AACAAATTA	AAAGTTGGTT	840
15	TTTTGAAAG	ATAAACAAAA	TTGACAAACC	TTTGCCAGAG	CTAAGAAAAA	AGGAAGAAAG	ACCTAAATAA	910
	ATAAAGTCAG	AGATGAAAAA	AGAGACATTA	CAACTGATAC	CACAGAAATT	CAAAGGATCA	CTAGAGGCTA	980
	CTATGAGCAA	CTGTACACTA	ATAAATTGAA	AAACCTAGAA	AAAATAGATA	AATTCCTAGA	TGCATACAC	1050
	CTACCAAGAT	TGAACCATGA	AGAAATCCAA	AGCCCAACAA	GACCAATAAC	AATAATGGGA	TATAAGCCAT	1120
20	AATAAAAAAGT	CTCTTAGCAA	AGAGAAGCCC	AGGACCCAAT	GGCTTCCCTG	CTGGATTTTA	CCAATCATTT	1190
	AAAGAAGAAT	GAATTTCCAAT	CCTACTCAAA	CTATTCTGAA	AAATAGAGGA	AAGAATACTT	CCAAACTCAT	1260
	TCTCATGGCC	CAGTATTACC	CTGATTCCAA	AACCAGACAA	AAACACATCA	AAAACAAACA	AACAAAAAAA	1330
	CAGAAAGAAA	GAAACTTACA	GGCCAAATATC	CCTGATGAAT	ACTGATACAA	AAATCCTCAA	CAAAACACTA	1400
	GCAAACCAAA	TTAAACAACA	CAATTCGAAAG	ATCATTCAAT	GTGATCAAGT	GGGATTTATT	CCAGGGATGG	1470
25	AAGGATGGTT	CAACATATGC	AAATCAATCA	ATGTGATACA	TCATCCCAAC	AAAATGAAGT	CAAAAAACTA	1540
	TATGATTATT	TCACTTTATG	CAGAAAAAGC	ATTTGATAAA	ATTCTGCACC	CTTCATGATA	AAAAACCTCA	1610
	AAAAACCAAG	TATACAAGAA	ACATACAGGC	CAGGCACAGT	GGCTCACACC	TGCGATCCCA	GCACCTCTGG	1680
	AGGCCAAGGT	GGGATGATTG	CTTGGGCCCA	GGAGTTTGAG	ACTAGCCTGG	GCAACAAAAT	GAGACCTGGT	1750
	CTACAAAAAA	CTTTTTTAAA	AAATTAGCCA	GGCATGATGG	CATATGCCCTG	TAGTCCAGC	TAGTCTGGAG	1820
30	GCTGAGGTGG	GAGAATCACT	TAAGCCTAGG	AGGTCGAGGC	TGCAGTGAGC	CATGAACATG	TCACCTGACT	1890
	CCAGCCTAGA	CAACAGAAC	AGACCCCACT	GAATAAGAGG	AAGGAGAAGG	AGAAAGGAGA	AGGGAGGGAG	1960
	AAGGAGGAGG	GAGGAGAAGG	AGGAGGTGGA	GGAGAAGTGG	AAGGGGAAGG	GGAAAGGAAA	AGGGAAGGAG	2030
	AAGAAACATA	TTTCAACATA	ATAAAGCCCC	TATATGACAG	ACCGAGGTAG	TATTATGAGG	AAAAACTGAA	2100
	AGCCTTTCCCT	CTAAGATCTG	GAAATGACA	AGGGCCCACT	TTCAACCACTG	TGATTCAACA	TAGTACTGAA	2170
	AGTCCTAGCT	AGAGCAATCA	GATAAGAGAA	AGAAATAAAA	GGCATCCAAA	CTGGAAGGGA	AGAAGTCAAA	2240
35	TTATCCTGTT	TGCAGATGAT	ATGATCTTAT	ATCTGGAATA	GACTTAAGAC	ACCACTAAAA	AACATTATGA	2310
	GCTGAAATTT	GGTACAGCAG	GATACAAAAA	CAATGTACAA	AAATCAGTAG	TATTTCTATA	TTCCAACAGC	2380
	AAACAATCTG	AAAAAGAAAC	CAAAAAAGCA	GCTACAAATA	AAATTAAACA	GCTAGGAATT	AACCAAGAAA	2450
	GTGAAAGATC	TCTACAATGA	AAACTATAAA	ATGTTGATAA	AAGAAATTGA	AGAGGGCACA	AAAAAGAAAA	2520
	AGATATTCCA	TGTTTATAGA	TTGGAAGAAAT	AAATCTGTAT	AAATGTGCTA	TACTACCCAA	AGCAATTATC	2590
40	AAATTTCAATG	CAATCCCTAT	TAAATACTA	ATGACGTTCT	TCACAGAAAT	AGAAGAAACA	ATTCTAAGAT	2660
	TTGTACAGAA	CCACAAAAGA	CCCAGAATAG	CCAAAGCTAT	CCTGACCAAA	AAGAACAAAA	CTGGAAGCAT	2730
	CACATTACCT	GACTTCAAAAT	TATACTACAA	AGCTATAGTA	ACCCAAACTA	CATGGTACTG	GCATAAAAAC	2800
	AGATGAGACA	TGGACCAGAG	GAACAGAATA	GAGAAATCCG	AAACAAATCC	ATGCATCTAC	AGTGAACCTCA	2870
45	TTTTTGACAA	AGGTGCCAAG	AACATACTTT	GGGGAAAGAA	TAATCTCTTC	AATAAATGGT	GCTGGAGGAA	2940
	CTGGATATCC	ATATGCAAAA	TACAATACT	AGAACTCTGT	CTCTCCAGAT	ATACAAAAGC	AAATCAAAAT	3010
	GGATGAAAGG	CTTAAATCTA	AAACCTCAAA	CTTTGCAACT	ACTAAAGAAA	AACACCGGAG	AACTCTTCCA	3080
	GGACATTGGA	GTGGGCAAG	ACTTCTTGAG	TAATTCCTCG	CAGGCACAGG	CAACCAAGGC	AAACACAGAC	3150
	AAATGGGATC	ATATCAAGTT	AAAAAGCTTC	TGCCAGCAAA	AGGAACAAT	CAACAAAGAG	AAGAGACAAC	3220
50	CCACAGAATG	GGAGAATATA	TTTGCAAACT	ATTCATCTAA	CAAGGAATTA	ATAACCACTA	TATATAAGGA	3290
	GCTCAAACTA	CTCTATAAGA	AAAAACACCTA	ATAAGCTGAT	TTTCAAAAAAT	AAGCAAAAGA	TCTGGGTAGA	3360
	CATTCTCTCA	AATAAGTCAT	ACAAATGGCA	AACAGGCATC	TGAAAATGTG	CTCAACACCA	CTGATCATCA	3430
	GAGAAATGCA	AATCAAAACT	ACTATGAGAG	ATCATCTCAT	CCCAGTTAAA	ATGGCTTTTA	TTCAAAAGAC	3500
	AGGCAATAAC	AAATGCCAGT	GAGGATGTGG	ATAAAGGAAA	ACCCTTGAC	ACTGTTGGTG	GGAAATGAAA	3570
55	TTGCTACCAC	TATGGAGAAC	AGTTTGAAAG	TTCTCAAAA	AACTAAAAAT	AAAGCTACCA	TACAGCAATC	3640
	CCATTGCTAG	GTATATACTC	CAAAAAAGGG	AATCAGTGTA	TCAACAAGCT	ATCTCCACTC	CCACATTTC	3710
	TGCAGCACTG	TTCATAGCAG	CCAAGGTTTG	GAAGCAACCT	CAGTGTTCCAT	CAACAGACGA	ATGGAAAAAG	3780
	AAAATGTGGT	GCACATACAC	AATGGAGTAC	TACGCAGCCA	TAAAAAAGAA	TGAGATCCTG	TCAGTTGCAA	3850
	CAGCATGGGG	GGCACTGGTC	AGTATGTTAA	GTGAAATAAG	CCAGGCACAG	AAAGACAAAC	TTTTTATGTT	3920
60	CTCCCTTACT	TGTGGGAGCA	AAAAATTAATA	CAATTGACAT	AGAAATAGAG	GAGAATGGTG	GTTCATAGAGG	3990
	GGTGGGGGAC	AGGGTGACTA	GAGTCAACAA	TAATTTATTG	TATGTTTAA	AATAACTAAA	AGAGTATAAT	4060
	TGGGTTGTTT	GTAACACAAA	GAAAGGATAA	ATGCTTGAAG	GTGACAGATA	CCCCATTTAC	CCTGATGTGA	4130
	TTATTACACA	TTGTATGCCCT	GTATCAAAAT	ATCTCATGTA	TGCTATAGAT	ATAAACCCCTA	CTATATTAAA	4200
	AATTAATAAT	TTAATGGCCA	GGCACGGTGG	CTCATGTCCG	TAATCCCAGC	ACTTTGGGAG	GGCCAGGCGG	4270
65	GTGGATCAC	TGAGGTCAAG	AGTTTGAAAC	CAGTCTGGCC	ACCATGATGA	AACCTGTCT	CTACTAAAGA	4340
	TACAAAAATT	AGCCAGGCGT	GGTGGCACAT	ACCTGTAGTC	CCAATCTACT	AGGAGGCTGA	GACAGGAGAA	4410
	TTGCTTGAAC	CTGGGAGGCG	GAGGTTGCAG	TGAGCCGAGA	TCATGCCACT	GCAGTGCAGC	CTGGGTGACA	4480
	GAGCAAGACT	CCATCTCAAA	ACAAAAACAA	AAAAAGAAAG	ATTAAAAATTG	TAATTTTAT	GTACCGTATA	4550
	AAATATATCT	CTACTATATT	AGAGTTTAAA	AATTAACA	ATTATAAAG	GTAATTAAAC	ACCTTAATCTA	4620
70	AAATAGAAC	AATGTATGTG	AGGTTTCTAG	CTTCTGAAGA	AGTAAAGTT	ATGGCCACGA	TGGCAGAAAT	4690
	GTGAGGAGGG	AACAGTGGA	GTTACTGTTG	TTAGACGCTC	ATACTCTCTG	TAAGTGACTT	AATTTTAACC	4760
	AAAGACAGGC	TGGGAGAAGT	TAAAGAGGCA	TTCTATAAGC	CCTAAAAACAA	CTGCTAATAA	TTGTGAAAGG	4830
	TAATCTCTAT	TAATTTACCA	TAATTACAGA	TATCTCTAAA	ATCGAGCTGC	AGAATTGGCA	CTGCTGATCA	4900
	CACCGTCTCT	TCATTACAGG	TGCTTTTTTT	TTTGTGTGCT	TGGAGATTTT	CGATTGTGTG	TTCGTGTTTG	4970
75	GTTAAACCTA	ATCTGTATGA	ATCCTGAAAC	GAAAAATGGT	GGTGATTTC	TCCAGAGAA	TTAGAGTACC	5040
	TGGCAGGAAG	CAGGTGGCTC	TGTGGACCTG	AGCCACTTCA	ATCTTCAAGG	GTCTCTGGCC	AAGACCCAGG	5110

	TCGAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAAGCTC	GGATGGGAAG	GGGCGATGAG	AAGCCTGCCT	5180
	CGTTGGTGAG	CAGCGCATGA	AGTGCCCTTA	TTTACGCTTT	GCAAAGATTG	CTCTGGATAC	CATCTGGAAA	5250
	AGGCGGCCAG	CGGGAATGCA	AGGAGTCAGA	AGCCTCCTGC	TCAAACCCAG	GCCAGCAGCT	ATGGCGCCCA	5320
5	CCCGGGCGTG	TGCCAGAGGG	AGAGGAGTCA	AGGCACCTCG	AAGTATGGCT	TAAATCTTTT	TTTCACTGA	5390
	AGCAGTGACC	AAGGTGTATT	CTGAGGGGAG	CTTGAGTTAG	GTGCCTTCTT	TAAACAGAA	AGTCATGGAA	5460
	GCACCTCTCT	CAAGGGAAAA	CCAGACGCCC	GCTCTGCGGT	CATTACCTC	TTTCTCTCT	CCCTCTCTTG	5530
	CCCTCGCGGT	TTCTGATCGG	GACAGAGTGA	CCCCCTGGA	GCTTCTCCGA	GCCCGTGCTG	AGGACCTCT	5600
	TGCAAGGGG	TCCACAGACC	CCGCGCTGG	AGAGAGGAGT	CTGAGCCTGG	CTTAATAACA	AACCTGGATG	5670
10	TGGCTGGGGG	CGGACAGCGA	CGGCGGAT	CAAAGACTTA	ATTCCATGAG	TAAATTCAAC	CTTCCACAT	5740
	CCGAATGGAT	TTGGATTTTA	TCTTAATATT	TTCTTAAATT	TCATCAAATA	ACATTACAGGA	CTGCAGAAAT	5810
	CCAAAGCGGT	AAAACAGGAA	CTGAGCTATG	TTTGCCAAAG	TCCAAGGACT	TAATAACCAT	GTTCAGAGGG	5880
	ATTTTTCGCC	CTAAGTACTT	TTTATTGGTT	TTTCAAGGT	GGCTTAGGGT	GCAAGGGGAA	GTACACGAGG	5950
	AGAGGCTG	GCGGCAGGGC	TATGAGCAGG	GCAGGGCCAC	CGGGGAGAGA	GTCCCGGGCC	TGGGAGGCTG	6020
15	ACAGCAGGAC	CACTGACCGT	CCTCCTGGG	AGCTGCCACA	TTGGGCAACG	GGAAGGCGGC	CACGCTCGCT	6090
	GTGACTCAGG	ACCCATACCC	GGCTTCTGG	GGCCACCCAC	ACTAACCCAG	GAAGTCACGG	AGCTCTGAAC	6160
	CCGTGGAAAC	GAACATGACC	CTTGCTGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAAATGAA	GTGGTGTGCA	6230
	GGAAATGGCC	ATGTAAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCCTTC	CATCTATTAT	CATCTTCACC	6300
	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAAACTCAG	TACAAACACC	6370
20	ACTCTTTTAC	TAGGCCACACA	GAGCAGCGSC	CACACCCCTG	ATATATTAAG	AGTCCAGGAG	AGATGAGGCT	6440
	GCTTTTCAGC	ACCAGGCTGG	GGTGACAACA	CGCGCTGAAC	AGTCTGTTC	TCTAGACTAG	TAGACCTGG	6510
	CAGGCACTCC	CCCAGATTCT	AGGGCTGTGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	GAGACTCAGC	6580
	CTGGGTGGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCTT	CCGCTCCAG	GGCTCAGCTT	CTCCAGCAGC	6650
	TTCTTAAACC	CTGGGTGGGC	CGTGTTCAG	CGCTACTGTC	TCACCTGTTC	CACCTGTGCT	CTCTCAGCG	6720
25	ACGTAGCTCG	CACGGTTCTT	CCTCACATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCACAT	CGGTGGAAGG	6790
	GAGGAGATTG	TGCGCCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	GTGACAGGTT	6860
	CTTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TTCCAGGCGC	TCCCGCTCTC	CTGTCTATCT	CCGGGCGCTG	6930
	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGTGTG	TCTCTGCCCG	CTAGGGTCTC	7000
	GGGGTTTTTA	TAGGCATAGG	ACGGGGCGGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAA	CATTGTGGTG	7070
30	TGAAAGTAGG	AGTGCCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGCCTGGG	GATGGAGCCC	CCGCCAGGGA	7140
	CCCGCCCTCT	TCTGCCACGC	ACTTTCCTCG	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	TTCCACAAGC	7210
	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GGCCACACGC	CCTGGGAATT	7280
	CACGCTACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCGG	CTGTTTTATT	TAAATAGCTT	7350
	CAAGAGGAGG	AAATCCCTGC	TAAATGTCC	TTTAAACAAAC	TGGTTAAACA	AACGGGTCCA	TCCGCACGGT	7420
35	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	GAATTACGCT	7490
	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAAG	AAAGAAATTC	ACCCATGGC	7560
	AGGGGAGTGG	TTAGGGGGGT	TAAAGACGGT	GGGGCGGGCA	GCTGGGGGCT	ACTGCACGCA	CCTTTTACTA	7630
	AAGCCAGTTT	CCTGGTTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	TGGGGATGGG	7700
	GGAAACCCGA	GGCTGTGCCA	TCCTTGGCAT	GCCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	AGAGATGCCC	7770
40	ACGTCCTGAT	TCCCCCAAAC	CTGTGGACAG	AACCCGCCCC	GCCCCAGGGC	CTTTGCAAGT	GTGATCTCCG	7840
	TAGGAGCCCT	GAGGTCTGGG	ATCCTCTGGG	ACTACCTGCA	GGCCCGAAAA	GTAATCCAGG	GGTCTGGGA	7910
	AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTACGTCTGA	GGCTGAAAAG	7980
	GGAGGGAGGG	CCTCGAGCGC	AGGCCCTGCA	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	GGGACCCCTC	8050
	ACGGAGCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	TCCGGCCTCC	8120
45	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTCAGAAA	GCAACAGGAA	8190
	ACCCATGAC	TGTGAATCTA	GGATTATTTT	AAACAAAGG	TTTACAGAAA	CATCCAAGGA	CAGGGCTGAA	8260
	GTGCCCTCCG	GCAAGGGCAG	GGCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	TACTTACTTT	8330
	CTGAGACAGA	GTTATGCTCT	TGTTGCCAG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	CTGCAACCTC	8400
	CGTCTCCTGG	GTTCAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTACAGC	GTGCACCAAC	8470
50	ACACCCGGCT	AATTTTGTAT	TTTATAGTAG	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	TTCAAGAACT	8540
	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	CACCTGCACCT	8610
	GGCCTATTTA	ACCATTTTAA	AACTTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	CATGGAGGCT	8680
	AATTTCCCTT	TTACTCAGGA	GTTACCCCTC	TTTGATATTT	TCTGTAATTC	TTCTAGACT	GGGGATACAC	8750
	CGTCTCTTGA	CATATTCACA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCAGCTCAG	GGGCAGCTGG	8820
55	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	ATCAGGGCGC	8890
	AAGTGTGGAC	ACTGTCTCTG	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	AAGTCCATCC	8960
	CTCCTACTCT	ACTGGGATTG	AGCCCCCTCC	CTATCCCCCC	CCAGGGGCGAG	AGGAGTTCCCT	CTCACTCCTG	9030
	TGGAGGAAGG	AATGATACTT	TGTTATTTTT	CACCTGCTGT	ACTGAATCCA	CTGTTTCATT	TGTTGGTTTG	9100
	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTTGCTCA	GGCTGGAGGG	AGTGCAATGG	CGCGATCTTG	9170
60	GCTTACTGCA	GCCTCTGCCT	CCCAGGTCCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	GCTGGGATTA	9240
	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGTATTTT	TAGTAGAGAC	GGGGGTGGGT	GGGGTTACCC	9310
	ATGTTTGGCA	GGCTGGTCTC	GAACCTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCTCTC	AAAGTGTCTG	9380
	GATTACAGGT	GTGAGCCACC	ATGCCCCAGCT	CAGAATTTAC	TCTGTTTAGA	AACATCTGGG	TCTGAGGTAG	9450
	GAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAGCCAA	TGATAGAATT	TTTTTATTGT	TGTTAGAACA	9520
65	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	TGCACCCATA	9590
	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	CCATGCACAT	9660
	GGTGTAAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTCTT	TCTCTTCCCT	CTTTTAAAT	TGTTGTTTCT	9730
	ATGTTGGCTT	CTCTGCAGAG	AACCACTGTA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	TTTCCAAACC	9800
	GCCCCCTTGC	CCTAGTGGCA	GAGACAATTC	ACAAACACAG	CCCTTTAAAA	AGGCTTAGGG	ATCACTAAGG	9870
70	GGATTCTTAG	AAGAGCGACC	TGTAATCTTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	GAGCGTGACA	9940
	CCCCAGGGAG	GGTCCGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCTCTC	GGCAGTTTCT	10010
	GAAAGTAGGA	AAGGTATCAT	TTAAGGTTGC	GTTTGTTAGC	ATTTCAGTGT	TTGCCGACCT	CAGCTACAGC	10080
	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AAACTTGAGC	AACCCGAGT	10150
	CTGGATTCTT	GGGAAGTCTT	CAGCTGTCTT	CGGGTTGTGC	CGGGGCCCCA	GCTCTGGAGG	GGACCACTGG	10220
75	CCGTGTGGCT	TCTACTGCTG	GGCTGGAGGT	CGGGCTCTCT	AGCTCTGACG	TCCGAGGCTT	GGAGCCAGGT	10290
	CGCTGGACCC	CGAGGCTGCC	CTGCCACCTG	TGCGGGCGGG	ATGTGACACG	ATGTTGGCTC	CATCTGCCAG	10360
	ACAGAGTGCC	GGGGCCACAG	GTCAAGGCCG	TTGTGGCTGG	TGTGAGGCGC	CCGGTGGCGG	GCCAGCAGGA	10430
	GCGCTTGCT	CCATTTCCCA	CCCTTCTCG	ACGGGACCCG	CCCGGTGGGT	GATTAACAGA	TTTGGGGTGG	10500

	TTTGCTCATG	GTGGGGACCC	CTCGCCGCT	GAGAACCTGC	AAAGAGAAAT	GACGGGCTG	TGTCAAGGAG	10570
	CCCAAGTCGC	GGGGAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCCGTCCAG	GGAGCAATGC	10640
	GTCCTCGGGT	TGCTCCCCAG	CCGCGTCTAC	GCGCTCCGT	CCTCCCTTC	ACGTCCGGCA	TTCTGGTGC	10710
5	CCGGAGCCCG	ACGCCCCCGG	TCCGGACCTG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	CGGGCCAAAG	10780
	GGTGCCCGCA	CGCACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCCCTCC	CGGGTTACCC	CACAGCCTAG	10850
	GCCGATTCTGA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCCT	GGGAGCGCGA	CGGGCGCGCG	10920
	GGCGGGGAAG	CGCGGGCCAG	ACCCCCGGGT	CCGCCCCGAG	CAGCTGCGCT	GTCGGGGCCA	GGCCGGGCTC	10990
	CCAGTGGGATT	CGCGGGCACA	GACGCCCAGG	ACC CGCTCC	CCACGTGGCG	GAGGGACTGG	GACCCGGGCG	11060
10	ACCCGTCTGT	CCCTTTCACC	TTCACGTCTC	GCCTCTCCG	CGCGGACCCC	GCCCCGTCCC	GACCCCTCCC	11130
	GGGTCCCCGG	CCCAGCCCCC	TCCGGGCCCT	CCCAGCCCCC	CCCTTTCCTT	TCCGGGGCCC	CGCCCTCTCC	11200
	TCCGGGGCGG	AGTTTCAGCG	AGCGCTGCGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	GGCCACCCCC	11270
	GCGATTGCCG	GCGCTCCCGG	CTGCCGAGCC	GTGCGCTCCC	TGCTGCGCAG	CCACTACCGC	GAGGTGCTGC	11340
	CGCTGGCCAC	GTTCGTGCGG	CGCCTGGGGC	CCCAGGGCTG	CGCGCTGGTG	CAGCGCGGGG	ACCCGGCGGC	11410
15	TTTCCCGCGG	CTGGTGGCCC	AGTGCTGGT	GTGCGTGCCC	TGGGACGCAC	GGCCGCCCCC	CGCCGCCCCC	11480
	TCCTTCCGCG	AGGTGGGCCCT	CCCCGGGCTC	GGCGTCCGGC	TGGGGTTGAG	GGCGGCCGGG	GGGAACCAGC	11550
	GACATGCGGA	GAGCAGCGCA	GGCGACTCAG	GGCGCTTCCC	CGCGAGGTGT	CTTGCTGAA	GAGAGTGGTG	11620
	GCCCGAGTGC	TGCAGAGGCT	GTGCGAGCGC	GGCGCAAGA	ACGTGCTGGC	CTTCGGCTTC	CGCTGCTGAG	11690
	ACGGGGCCCG	CGGGGGCCCC	CCCAGGCCCT	TCACCAACAG	CGTGCGCAGC	TACCTGCCCA	ACACGGTGAC	11760
20	CGACGCACTG	CGGGGGAGCG	GGGCGTGGGG	GCTGCTGCTG	CGCCCGCTGG	GCGACGACGT	GCTGGTTTAC	11830
	CTGCTGGCAC	GCTGCGCGCT	CTTTGTGCTG	GTGGCTCCCC	GCTGCGCTTA	CCAGGTGTGC	GGCCCGCGCG	11900
	TGTACGAGCT	CGGCGCTGCC	ACTCAGGCCC	GGCCCCGCC	ACACGCTAGT	GGACCCCGAA	GGCGTCTGGG	11970
	ATGCGAACGG	GCCTGGAACC	ATAGCGTCAG	GGAGGCCGGG	GTCCCCCTGG	GCCTGCCAGC	CCCCGGTGGC	12040
	AGGAGGCCGG	GGGGCAGTGC	ATAGCCGAAGT	CTGCCGTTCG	CCAAGAGGCC	CAGGCGTGGC	GCTGCCCTGT	12110
25	AGCCGGAGCG	GACGCCCTGT	GGGCGAGGGT	CCTGGGCCCC	CCCGGGCAGG	ACCGTGGGAC	CGAGTGACCG	12180
	TGGTTTCTGT	GTGGTGTAC	CTGCCAGACC	CGCCGAAGAA	GCCACCTCTT	TGGAGGGTGC	GCTCTGTGGC	12250
	ACGGGCCACT	CCCACCCATC	CGTGGGCGCG	CAGCACCAGC	CAGGCCCCCC	ATCCACATCG	GGCCACCACG	12320
	GTCCCTGGGA	CACGCCCTGT	CCCCCGGTGT	ACGCCGAGAC	CAAGCACTTC	CTCTACTCCT	CAGCGCACAA	12390
	GGAGCAGCTG	CGGCCCTCCT	TCCTACTCAG	CTCTCTGAGG	CCCAGCCTGA	CTGGCGCTCG	GAGGCTCGTG	12460
30	GAGACCATCT	TTCTGGGTTT	CAGGCCCTGG	ATGCCAGGGA	CTCCCGCAG	GTTGCCCGCG	CTGCCCGTCT	12530
	GCTACTGGCA	AATGCGGCC	CTGTTTCTGG	AGCTGCTTGG	GAACACGCG	CAGTGCCCTT	ACGGGGTGTCT	12600
	CCTCAAGACG	CACTGCCCGG	TGCGAGCTGC	GGTCACCCCA	GCAGCCGGTG	TCTGTGCCCG	GGGAAGGCC	12670
	CAGGGCTCTG	TGGCGGCCCG	CAGGAGGAGG	GACACAGACC	CCCGTCCGCT	GSTGAGCTG	CTGCCCGAGC	12740
	ACAGCAGGCC	CTGGCAGGTG	TACGGCTTCG	TGCGGGCGCT	CCTGCGCCGG	CTGGTGCCCC	CAGGCCCTCT	12810
35	GGGCTCCAGG	CACAACGAAC	GCCGCTTCTT	CAGGAACACC	AAGAAGTTCA	TCTCCCTGGG	GAAGCATGCC	12880
	AAGCTCTCGC	TGCAGGAGCT	GACGTGGAAG	ATGAGCGTGC	GGGACTGCGC	TTGGCTGCGC	AGGAGCCGAG	12950
	GTGAGGAGGT	GGTGGGCCGT	GAGGGGCCAG	GCCCCAGAGC	TGAATGCAGT	AGGGGCTCAG	AAAAGGGGCG	13020
	AGGCAGAGCC	CTGGTCTCCT	GTCACGTGGG	CACACGTGGC	TTTTTCGCTA	GGACGTGCGG	CAGGCTGAG	13090
	TGGACACGGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGCATAA	ACTTACGAGG	TTCACTTCA	13160
40	CGTTTTGATG	GACACCGCGT	TTCAGGCGCC	CGAGGCCAGA	GCAGTGAACA	GAGGAGGCTG	GGCCGGCAG	13230
	TGGAGCCGGG	TTGCCGGCAA	TGGGGAGAAG	TGCTGGAAG	CACAGACGCT	CTGCCGAGGG	TGCTTGCAGG	13300
	TTACTATATA	TCCTCTTCGC	AATTTCAAGG	GTGGGAATGA	GAGGTGGGGA	CGAGAACCCC	CTCTTCTCTG	13370
	GGGTGGGAGG	TAAGGGTTTT	GCAGGTGCAC	GTGGTCAGCC	AATATGCAGG	TTTGTGTTTA	AGATTTAATT	13440
	GTGTGTGAC	GGCCAGGTGC	GGTGGCTCAC	GCCGGTAATC	CCAGCACTTT	GGGAAGCTGA	GGCAGGTGGA	13510
45	TCACCTGAGG	TACAGAGTTT	GAGACCAAGC	TGACCAACAT	GGTGAACCCC	TATCTGTACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATGGTGG	TGTGTGCTCG	TAATCCAGC	TACTTGGGAG	GCTGAGGCAG	GAGAACTACT	13650
	TGAACCCAGG	AGGCGGAGGC	TGCAATGAGC	TGAGATTGTG	CCATTGTACT	CCAGCCTGGG	GCACAAGAGT	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGT	CGTTGATTGT	GCCAGGACAG	GCTAGAGGGA	GGGAGATAAG	13790
	ACTGTTCTCC	AGCACAGATC	CTGGTCCCAT	CTTTAGGTAT	GAAGAGGGCC	ACATGGGAGC	AGAGGACAGC	13860
50	AGATGGCTCC	ACCTGCTGAG	GAAGGGACAG	TGTTTGTGGG	TGTTACAGGG	ATGGTGCTCG	TGGCCCTGCG	13930
	CGTGTCGCCA	CCCTGTTTTT	CTGGATTGGA	TGTTGAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGTCCCGAG	14000
	TGCTCCCGAG	CCCTACCGTG	GCAGCTAGAA	GAAGTCCCGA	TTTCAACCCC	TCCCAACAAA	CTCCCAAGAC	14070
	ATGTAAGACT	TCCGGCCATG	CAGACAAGGA	GGGTGACCTT	CTTGGGGCTC	TTTTTTTCTT	TTTTTCTTTT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACTCC	TAACACCGTT	TTCTGTGTAC	AGTGCAGAA	14210
55	TGCTAACTCG	CGCGTGTGTA	CAGCAGGTTC	CTTGAAATGC	TGCGTCTTGC	GTGACTGGAA	GTCCCTACCC	14280
	ATCGAACGGC	AGCTGCCTCA	CACCTGCTGC	GGCTCAGGTG	GACCACGCCG	AGTCAGATAA	CGCTCATGCA	14350
	ACCCAGTTTT	GCTTTTTGTG	CTCCAGCTTC	CTTCGTTGAG	GAGAGTTTGA	GTTCTCTGAT	CAGGACTCTG	14420
	CCTGTCACTG	CTGTTCTCTG	ACTTCAGATG	AGGTCACAAT	CTGCCCTTGG	CTTATGCAGG	GAGTGAGGCG	14490
	TGGTCCCCGG	GTGTCCCTGT	CACGTGCAAG	GTGAGTGAGG	CGTTGCCCCC	AGGTGTCCCT	GTACGCTGTA	14560
60	GGGTGAGTGA	GGCGCGGCC	CCGGGTGTCC	CTGTCCCGTG	CAGCGTGATT	GAGGTGTGGC	CCCCGGGTGT	14630
	CCCTGTACCG	TGTAGGGTGA	GTGAGCGGCC	ATCCCCGGGT	GTCCCTGTCA	CGTGTAGGCT	GAGTGAGGCG	14700
	TGGTCCCCGG	GTGTCCCTGT	CCCGTGCAGG	GTGAGTGAGG	CACGTGCCCC	GGGTGTCCCT	GTACGCTGCA	14770
	GGGTGAGTGA	GGCGCGGTCC	CCGGGTGTCC	CTCTCAGGTG	TAGGGTGAGT	GAGGCGCGGC	CCCAGGGTGT	14840
	CCCTGTACCG	TGTAGGGTGA	GTGAGGCACC	GTCCCTGGGT	GTCCCTCCCA	GGTATAGGGT	GAGTGAGGCA	14910
65	CTGTCCCCGG	GTGTCCCTGT	CACGTGCAAG	GTGAGTGAGG	CGCGGCCCCC	GGGTGTCCCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGGTGTCC	CTGGGTGTCC	CTGTCTCGTG	TAGGGTGAGT	GAGGCTCTGT	CCCCAGGTGT	15050
	CCTTGGCGTT	TGCTCACTTG	AGCTTGCTCC	TGAATGTTTG	CTCTTTCTAT	AGCCACAGCT	GGCCGGGTGT	15120
	CCCAATTGCT	GGGTAGATGG	TGCAGGCACA	GTGCTGGTCC	CCAAGCTAT	CTTTTCTGAT	GCTCGGCTCT	15190
	TCTTGGTCAC	CTCTCCGTTT	CATTTTGCTA	CGGGGACACG	GGACTGCAGG	CTCTCGCCTC	CCCGTGCCCA	15260
70	GGCACTGCAG	CCACAGCTTC	AGGTCCGCTT	GCCTCTGTTG	GGCTGGGCTT	GCTCACCACG	TGCCCGCCAC	15330
	ATGCATGCTG	CCAATACTCC	TCTCCAGCT	TGTCTCATGC	CGAGGCTGGA	CTCTGGGCTG	CCTGTGTCTG	15400
	CTGCCACGTG	TTGCTGGAGA	CATCCAGAA	AGGGTTCTCT	GTGCCCTGAA	GGAAAGCAAG	TCACCCAGC	15470
	CCCTCACTTT	GTCTGTGTTT	CTCCCAAGCT	GCCTCTGTC	TTGGCCCTCT	TGGGTGGGTG	GCAACGCTTG	15540
	TCACCTTATT	CTGGGCACCT	GGCGCTCATT	GCTTAGGCTG	GGCTCTGCGT	CCAGTCGCCC	CCTCACATGG	15610
	ATTACGCTCC	AGCCACAGGT	TGGAGTGTCT	CTGTCTGTCT	CCTGCTCTGA	GACCCACGTG	GAGGGCCGGT	15680
75	GCTTCCGCCA	GCCTTCGTGA	GACTTCCCTC	TGGGTCTTAA	GTTTGAATTT	TCACTGATTT	ACCTCTGACG	15750
	TTTCTATCTC	TCCATTGTAT	GCTTTTCTTT	GGTTTATCTT	TTCAATTCCTT	TTCTAGCTTC	TTAGTTTATG	15820
	CATGCGCTTC	CCTCTAAGTG	CTGCGTTTACC	TGCACCCCTG	GTTTTGATGT	GAAGTAATCT	CAACATCAGC	15890

	CACTTTCAAG	TGTTCTTAAA	ATACTTCAAA	GTGTTAATAC	TTCTTTTAAAG	TATTCTTATT	CTGTGATTTT	15960
	TTTCTTTTGTG	CACGCTGTGT	TTTGACGTGA	AATCATTTTG	ATATCAGTGA	CTTTTAAAGTA	TTCTTTAGCT	16030
	TATTCTGTGA	TTTCTTTGAG	CAGTGAGTTA	TTTGAACACT	GTTTATGTTC	AAGATATGTA	GAGTATCAAG	16100
	ATACGTAGAG	TATTTTAAAGT	TATCATTTTA	TTATTGATTT	CTAACTCAGT	TGTGTAGTGG	TCTGTATAAT	16170
5	ACCAATTATT	TGAAGTTTGC	GGAGCCTTGC	TTTGTGATCT	AGTGTGTGCA	TGGTTTCCAG	AACTGTCCAT	16240
	TGTAAATTTG	ACATCCTGTC	AATAGTGGGC	ATGCATGTTT	ACTATATCCA	GCTTATTAAG	GTCCAGTGCA	16310
	AAGCTTCTGT	CTCCTCTAG	ATGCATGAAA	TTCCAAGAAG	GAGGCCATAG	TCCCTCACCT	GGGGGATGGG	16380
	TCTGTTCATT	TCTTCTCGTT	TGGTAGCATT	TATGTGAGGC	ATTGTTAGGT	GCATGCACGT	GGTAGAATTT	16450
10	TTATCTTCCCT	GATGAGTGAA	TCTTTTGGAG	ACTTCTATGT	CTCTAGTAAT	CTAGTAATTC	TTTTTTTAAA	16520
	TTGCTCTTAG	TACTGCCACA	CTGGGCTTCT	TTTGATTAGT	ATTTTCTGTC	TGTGTCTGTT	TTCTGCCTTT	16590
	AAATTATATA	TATATATATA	TTTTTTTTTT	TTTTGAGACA	GAGTCTTGGT	CTGTCCGCCA	GGGTGAGTGC	16660
	AGTGGTGTGA	TCACAGGTCA	GTGTAACCTT	TACCTTCTGG	CCTGAGCCGT	CCTCTCACCT	CAGCCTCCTG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCGCTA	CACCTGGCTA	ATTTTAAAT	TTTTTCTGGA	GACAGGGTCT	16800
15	TGCTGTGTTG	CCCAGGCTGG	TCTCAAACCT	TTGGACTCAA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
	CTGAATTACA	GGCATGAGCC	ACCATGTCTG	GCCTAAATTT	CAACACTTTT	ATATTCCTAT	AGTGTGGGTA	16940
	TGTCTCTGTA	ACAGCATGTA	GGTGAAATTC	CAATCCAGTC	TGACAGTCGT	TGTTTAACTG	GATAACCTGT	17010
	TTTATTTTCA	TTTTTTTGTG	ACTAGAGACC	CGCCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTTGATGTT	17080
	CCTCGTTCCC	TTGTTTCTCA	CCACCTCTTG	GGTTGCCATG	TGCGTTTCTT	GCCGAGTGTG	TGTTGATCTT	17150
20	CTCGTTGCCCT	CCTGGTCACT	GGGCAATTTG	TTTTATTCTT	CTTTGCTTAG	TGTTAGCCCC	TGATCTTTTT	17220
	ATTGTGCTTG	TTTGCTTTG	TTTATTGAGA	CAGTCTCACT	CTGTCAACCA	GGCTGGAGTG	TAATGGCACA	17290
	ATCTCGGCTC	ACTGCAACCT	CTGCCCTCT	GGTTCAAGCA	GTTCCTATTG	CTCAACCTCA	TGAGTAGCTG	17360
	GGATTACAGG	CGCCACCCAC	CACGCTGGC	TAATTTTTGT	ATTTTATGTA	GAGATAGGCT	TTCACCATGT	17430
	TGGCCAGGCT	GGTCTCAAC	TCTGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACAG	TGCTGGGATT	17500
25	ACAGGTGCAA	GCCACCGTGC	CCGGCATACC	TTGATCTTTT	AAAAATGAAGT	CTGAAACATT	GCTACCCCTG	17570
	TCCTGAGCAA	TAAGACCCCT	AGTGTATTTT	AGCTCTGGCC	ACCCCCAGC	CTGTGTGCTG	TTTTCCCTGC	17640
	TGACTTAGTT	CTATCTCAGG	CATCTTGACA	CCCCCACAAG	CTAAGCATT	TAAATATTGT	TTTCCGTGTT	17710
	GAGTGTCTT	GTAGCTTTGC	CCCCGCCCTG	CTTTTCTCTC	TTTGTTCCTC	GTCTGTCTTC	TGCTCAGGCC	17780
	CGCCGCTCTG	GGGTCCCTTT	CTTGTCTCTT	TGCGTGGTTC	TTCTGTCTTG	TTATTGCTGG	TAAACCCAG	17850
30	CTTTACCTGT	GCTGGCCTCC	ATGGCATCTA	GCGACGCTCG	GGGACCTCTG	CTTATGATCG	ACAGATGAAG	17920
	ATGTGGAGAG	TCACAGGAGG	GGCGGTCATC	TGGCCCGGTG	AGTGTCTGGA	GCACACAGTG	GCCAGCGTTC	17990
	CTTAGCCAGT	GAGTGACAGC	AACGTCGCT	CGGCCCTGGT	TCAGCCTGGA	AAACCCAGG	CATGTGCGGG	18060
	TCTGGTGGCT	CCGCGGTGTC	GAGTTTGAAA	TCGCGCAAAC	CTGCGGTGTG	GCGCCAGCTC	TGACGGTGTCT	18130
	GCCTGGCGGG	GGAGTGTCTG	CTTCTCTCCT	TCTGCTTGGG	AACCAGGACA	AAGGATGAGG	CTCCGAGCCG	18200
35	TTGTGCGCCA	ACAGGAGCAT	GACGTGAGCC	ATGTGGATAA	TTTTAAATTT	TCTAGGCTGG	GCGCGGTGGC	18270
	TCACGCGCTG	AATCCACGA	CTTTGGGAGG	CCAAGGCGGG	TGGATCACGA	GGTCAGGAGG	TCGAGACCAT	18340
	CCTGGCCAA	ATGATGAAC	CCCATCTGTA	CTAATAACAC	AAAAATTAGC	TGGCGGTGGT	GGCGGGTGCC	18410
	TGTAATCCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATTG	CTTGAACCTG	GGAGTTGGAA	GTTCAGCTGA	18480
	GCCGACATTG	CACCACTGCA	CTCCAGCCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAAAA	18550
40	AAAAAATAAA	AATTTAGTA	GCCACATTAA	AAAAAGTAAA	AAGAAAAGGT	GAAATTAATG	TAATAATAGA	18620
	TTTTACTGAA	GCCAGCATG	TCCACACCTC	ATCATTTTAG	GGTGTATTG	GTGGAGCATG	CACCTCAGAG	18690
	ACATTTGACA	TTTTTTGAGC	TTTGTCTGCG	GGATCCCGTG	TGTAGGTCCC	GTGCGTGGCC	ATCTCGGCCCT	18760
	GGACCTGCTG	GGCTTCCCAT	GGCCATGGCT	GTGTACCCAG	ATGGGTGACG	TCCGGGATGA	GGTCGCCAGG	18830
	CCCTCAGTGA	GCTGGATGTG	CAGTGTCCGG	ATGGTGACG	TCTGGGATGA	GGTCGCCAGG	CCCTGCTGTG	18900
45	AGCTGGATGT	GTGGTGTCTG	GATGGTGACG	GTCAGGGGTG	AGGTCTCCAG	GCCCTCGGTG	AGCTGGAGGT	18970
	ATGGAGTCCG	GATGATGCAG	GTCGCGGGTG	AGGTGCGCAG	GCCCTGCTGT	GAGCTGGATG	TGTGGTGTCT	19040
	GGATGGTGCA	GGTCAGGGGT	GAGGTCTCCA	GGCCCTCGGT	AAGCTGGAGG	TATGGAGTCC	GGATGATGCA	19110
	GGTCCGGGGT	GAGGTGCGCA	GGCCCTGCTG	TGAGCTGGAT	GTGTGGTGTG	TGGATGGTGC	AGGCTCTGGG	19180
	TGAGGTCAAC	AGGCCCTGCG	GTGAGCTGGG	TGTGCGGTGT	CTGGATGGTG	CAGGTCTGGA	GTGAGGTCCG	19250
50	CAGACGGTGC	CAGACCATGC	GGTGAGCTGG	ATATGCGGTG	TCCGGATGGT	GCAGGTCTGG	GGTGAGGTGT	19320
	CCAGGCCCTG	CTGTGAGTTG	GATGTGGGGT	GTCCGGATGC	TGCAGGTCCG	GTGTGAGGTG	ACCAGGCCCT	19390
	GCTGTGAGCT	GGATGTGTGG	TGTCTGGATG	GTGCAAGTCT	GGGGTGAAGG	TCCGACGGCC	CTGCTTGTG	19460
	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GTCTGGAGTG	AGGTGCGCAG	GCCCTCGGTG	AGCTGGATGT	19530
	GCAGTGTCCA	GATGGTGCAG	GTCGCGGGTG	AGGTGCGCAG	ACCCTGCGGT	GAGCTGGATG	TCCGCTGTCT	19600
55	GGATGGTGCA	GGTCTGGAGT	GAGGTGCGCA	GGCCCTCGGT	GAGCTGGATG	TATGGAGTCC	GGATGGTGCC	19670
	GGTCCGGGGT	GAGGTGCGCA	GACCTGCTG	TGAGCTGGAT	GTGCGGTGTC	TGGATGGTAC	AGGTCTGGAG	19740
	TGAGGTGCGC	AGACCTGCTG	GTGAGCTGGA	TATGCGGTGT	CCGGATGGTG	CAGGTCAAGG	GTGAGGTCTC	19810
	CAGGCCCTCG	GTGAGCTGGA	GGTATGGAGT	CCGGATGATG	CAGGTCCCGG	GTGAGGTCCG	CAGGCCCTCG	19880
	TGTGAACCTG	ATGTGCGGGC	TCTGGATGGT	GCAGGTCTGG	GGTGTGGTCC	CCAGGCCCTC	GGTGAGCTGG	19950
60	AGGTATGGAG	TCCGGATGAT	GCAGGTCCGG	GGTGAGGTGC	CCAGGCCCTG	CTGTGAGCTG	GATGTGCGGC	20020
	GTCTGGATGG	TGCAGGTCTG	GGGTGTGGTC	GCCAGGCCCT	CGGTGAGCTG	GAGGTATGGA	GTCCGGATGA	20090
	TGCAGGTCCG	GGGTGAGGTT	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGCACTCCG	20160
	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGCAAGTCT	GGGTGAGGTC	20230
	CACCAGGCCCT	TGCGGTGAGC	TGGTTGTGCG	GTGTCCGGTT	GCTGCAAGTC	CGGGGTGAGT	TGCGAGGCCCT	20300
65	CTCGGTGAGC	TGGATGTGCG	GTGTCCCGCT	GTCCGGATGG	TGCAGGTCCA	GGGTGAGGTC	GCTAGGCCCT	20370
	TGGTGGGCTG	GATGTGCGCT	TGCGGATGG	GCGGTGAGTC	GGGTGAGGTC	GCCAGGCCCT	TGGTGAGCTG	20440
	GATGTGCGGT	GTCTGCATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGTGGT	20510
	TGCCGGATGG	TGCAGGTCCG	GCGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCGG	TGCTTGGATG	20580
	GTGCAAGTCC	GGGTGAGGTT	AGCCAAGGCC	TTCCGTGAGC	TGGATGTGGG	GTGTCCGGAT	GGTGCAAGTC	20650
70	CGGGGTGAGG	TGCGCAGGCC	CTGCGGTAG	CTGGATATGC	GGTGTCCGGA	TGGTGCAAGT	CCGGGGTGAG	20720
	GTCACCAAGC	CCTGCGGTTA	GCTGGATGTG	CGGTGTCTGG	ATGGTGCAAG	TCCGGGGTGA	GGTCCGCCAG	20790
	CCCTGCTGTG	AGCTGGATGT	GCTGTATCCG	GATGTGTGAG	GTCGGGGGTG	AGGTGCGCCG	GCCCTGCAAGT	20860
	GAGCTGGATG	TGCTGTATCC	GGATGGTGCA	GGTCTGGCGT	GAGGTGCGCA	GGCCCTGCGG	TTAGCTGGAT	20930
	ATGCGGTGTC	GGATGGTGCA	GGTCCGGGGT	GAGGTGCGCA	GGCCCTGCGG	TTAGCTGGAT	GTGCGGTGTC	21000
75	CGGATGGTGC	AGGTCTGGGG	TGAGGTGCGC	AGGCCCTGCT	GTGAGCTGGA	TGTGCTGTAT	CCGGATGGTG	21070
	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTGC	GGTGAGCTGG	ATGTGCTGTA	TCCGGATGGT	GCAAGGTCTG	21140
	CGTGAGGTCC	CCAGGCCCTG	CGGTGAGCTG	GATGTGCAAGT	GTACGGATGG	TGCAGGTCCG	GGGTGAGGTC	21210
	GCCAGGCCCT	GCGGTGGGCT	GTATGTGTGT	TGCTGAGATG	GTGCAAGTCC	GGGGTGAAGT	GCCAGGCCCT	21280

	TGCGGTGAGC	TGGATGTGTG	GTGTCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TCGCCAGGCC	CTCGGTGAGC	21350
	TGGATATGCG	GTGTCCCGCT	GTCCGAATGG	TGCAGGTCCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
	GATGTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGCCGT	21490
5	GTCCGGATGG	TGCAGGTCCG	GGGTGAGGTC	ACCAGGCCCT	CGGTGATCTG	GATGTGGCAT	GTCTTCTCG	21560
	TTAAGGGGT	TGGCTGTGTT	CGGCCCGCAG	AGCACCGTCT	CGGTGAGGAG	ATCCTGGCCA	AGTTCTTCCA	21630
	CTGGCTGATG	AGTGTGTACG	TCGTCAAGCT	GCTCAGGTCT	TTCTTTTATG	TCACGGAGAC	CACGTTTCAA	21700
	AAGAACAGGC	TCTTTTCTA	CGGGAAGAGT	GTCTGGAGCA	AGTTGCAAAAG	CATTGGAAATC	AGGTACTGTA	21770
	TCCCAACGCC	AGGCCTCTGC	TTCTCGAAGT	CCTGGAAACAC	CAGCCCGGCC	TCAGCATGCG	CTGTCTCTCA	21840
10	CTTGCTCTGT	CTTCCCTGGC	TGTGCAGCTC	TGGGTGGGA	GCCAGGGGCC	CCGTACACAG	CCTGGTCCAA	21910
	GTGGATTCTG	TGCAAGGCTC	TGACTGCCTG	GAGCTCAGCT	TCTCTTACTT	GTAATAATCAG	GAGTTTGTGC	21980
	CAAGTGGTCT	CTAGGGTTTG	TAAAGCAGAA	GGGATTTAAA	TTAGATGGAA	ACACTACCAC	TAGCCTCTCT	22050
	GCCTTTCCCT	GGGATGTGGG	TCTGATTCTC	TCTCTCTTTT	TTTTTCTTTT	TTTGAGATGG	AGTCTCACCT	22120
	TGTTGCCAG	GCTGGAGTGC	AGTGGCATAA	TCTTGGCTCA	CTGCAACCTC	CACCTCTGCT	GTTTAAAGCGA	22190
15	TTACCCAGCC	TCAGCCTCCT	AAGTAGCTGG	GATTACAGGC	ACCTGCCACC	ACGCTCTGGT	AATTTTGTGA	22260
	CTTTTAGAG	AGACGGGGTT	TCACCATGTT	GGCCAGGCTG	GTCTCGAACT	CATGACCTCA	GGTGATCCAC	22330
	CCACCTTGGC	CTCCCAAAGT	GCTGGGTTTA	CAGGCTAAGC	CACCGTGCCC	AGCCCCCGAT	TCTCTTTTAA	22400
	TTTCATGCTG	TCTGTATGAA	TCTTCAATCT	ATTGGATTAA	GGTCATGAGA	GGATAAAATC	CCACCCACTT	22470
	GGCGAGCTAC	TGCAGGGAGC	ACCTGTGCAG	GGGATAGGAG	AGTTCCACCA	AGTCTTAACCT	TGAGCTAATC	22540
20	TCTAGGTGGC	TGCATTTGAA	TGGCTGTGAG	ATTTTGTCTG	CAATGTTCCG	CTGATGAGAG	TGTGAGATTG	22610
	TGACAGATTG	AAGCTGGATT	TGCATCAGTG	AGGGACGGGA	GCGCTGGTCT	GGGAGATGCC	AGCCTGGGCTG	22680
	AGCCAGGCC	ATGGTATTAG	CTTCTCCGTG	TCCCGCCAG	GCTGACTGTG	GAGGGCTTTA	GTCAAGAAAT	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCACACTCG	AGTCCCTGGG	GGGCTTGTG	ACACCCCATG	CCCCAAATCA	22820
	GGATGTCTGC	AGAGGGAGCT	GGCAGCAGAC	CTCGTCAGAG	GTAACACAGC	CTCTGGGCTG	GGGACCCCGA	22890
25	CTGGTGTCTG	GGGCCATTTC	CTTGATCTGT	GGGGAGGGTC	AGGGCTTTCC	CTGTGGGAAC	AGTTAATATC	22960
	ACAATGACAC	TTACTTAGAC	TTTACACGTA	TTTAATGGTG	TGCGACCCAA	CATGGTCATT	TGACCAAGTAT	23030
	TTTGAAAGA	ATTTAATTGG	GGTGACCGGA	AGGAGCAGAC	AGACGTGGTG	GTCCCCAAGA	TGCTCCTTGT	23100
	CACTACTGGG	ACTGTTGTTT	TGCTTGGGGG	GCCTTGGAGG	CCCTCTCTCC	CTGGACAGGG	TACCGTGCTT	23170
	TTTCTACTCT	GCTGGGCTGT	CGGCTCGCGG	TCAGGGCACC	AGTCCCGGAG	CACCCGCGGC	CCCACTGTCC	23240
30	ACGGAGTGCC	AGGCTGTCTG	CCACAGATGC	CCAGGTCCAG	GTGTGGCCGC	TCCAGCCCCC	GTGCCCTTCT	23310
	GGGTGGTTT	GGGGGAAAG	GCCAGAGGCA	GAGGTGTCTG	GAGACTGGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCCTTGGC	TGAGCTGCCC	TGAGCAGCCT	CTCCCGCCCT	CTCCATCTGA	AGGGATGTGG	CTCTTCTACT	23450
	CTGGGGGTC	TGCTTGGGGG	CAGCCTTGGG	CTACCCCAAG	GGCTGTACCA	GAGGGACAGG	CATCCTGTGT	23520
	GGAGGGGAT	GGGTTCACGT	GGCCCCAGAT	GCAGCCTGGG	ACCAGGCTCC	CTGTGTCTGA	TGGTGGGACA	23590
35	GTACCCCTGG	GGGTTGACCC	CGGGACTGGG	CGTCCCCAGG	GTTGACTATA	GGACCAAGTG	TCCAGGTGCC	23660
	CTGCAAGTAG	AGGGGCTCTC	AGAGGCTCTC	GGCTGGCATG	GGTGGACGTG	GCCCGCGGCA	TGGCCTTCAG	23730
	CCTGTGCTGC	CGTGGGTGCC	CTGAGCCCTC	ACTGAGTCTG	TGGGGGCTTG	TGGCTTCCCG	TGAGCTTCCC	23800
	CTATGCTCTG	TGCTTGGCTG	AGCAAGCCTC	CTGAGGGGCT	CTCTATTGCA	GACAGCACTT	GAGGAGGGTG	23870
	AGCTGCGGG	AGCTGTCCGA	AGCAAGGCTC	AGGCAAGCAT	GGGAAGCCAG	GCCCGCCCTG	CTGACGTCCA	23940
40	GACTCCGCTT	CATCCCCAAG	CTGACGGGG	TGCGGCCGAT	TGTGAACATG	GACTACGTCT	TGGGAGCCAG	24010
	AACGTTCCCG	AGAGAAAAGA	GGGTGGCTGT	GCTTTGGTTT	AATCTCCTTT	TTAAACAGAA	GTGCGTTTGA	24080
	AGCCCACTT	TGGTATCAGC	TTAGATGAAG	GGCCCGGAGG	AGGGGCCACG	GGACACAGCC	AGGGCCATGG	24150
	CACGGGCCCA	ACCCATTGTG	GGCCACAGTG	AGGTGGCCGA	GGTGGCCGGT	CCTCCAGAAA	AGCAGCGTGG	24220
	GGGTGTAGGG	GGAGCTCCTG	GGGCAAGGAG	AGGCTCTGAG	GACCACAAGA	AGCAGCCGGG	CCAGGGGCTG	24290
45	GATGCAGCAC	GGCCCGAGGT	CCTGGATCCG	TGCTCTGCTG	TGGTGGCGAG	CTCCGCTGCG	CTTCCGCTTA	24360
	CGGGGGCCCG	GGACACAGGC	ACGACTGCCA	GGAGCCCAAC	GGGCTCTGAG	GATCCTGGAC	CTTGGCCCAT	24430
	GGCTCCTGCA	CCCCACCCCT	GTGGCTGCGG	TGGCTGCGGT	GACCCCGTCA	TCTGAGGAGA	GTGTGGGGTG	24500
	AGGTGGACAG	AGGTGTGGCA	TGAGGATCCC	GTGTGCAACA	CACATGCGGC	CAGGAACCCG	TTTCAACAG	24570
	GGTCTGAGGA	AGCTGGGAGG	GGTCTAGGTT	CCCCGGTCTG	GGTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24640
50	TCTCCCTGG	GTCCCTATGG	TGGGTGGGG	ACTTGGCCGG	ATCCACTTTC	CTGACTGTCT	CCCATGCTGT	24710
	CCCCGCCAGG	CCGAGCGTCT	CACCTCGAGG	GTGAAGGCAC	TGTTACGCGT	GCTCAACTAC	GAGCGGGCGC	24780
	GGCGCCCCGG	CCTCCTGGGC	GCCTCTGTGC	TGGGCCCTGGA	CGATATCCAC	AGGGCCTGGC	GCACCTTCGT	24850
	GCTGCGGTGT	CGGGCCCAAG	ACCCGCCGCC	TGAGCTGTAC	TTTGTCAAGG	TGGGTGGCCG	GGACCCCGCT	24920
55	GAGCAGCCCT	GCTGGACCTT	GGGAGTGGCT	GCCTGATTGG	CACCTCATGT	TGGGTGGAGG	AGGTACTCCT	24990
	GGGTGGGCCG	CAGGGAGTGC	AGGTGACCTT	GTCACTGTGT	AGGACACACC	TGGCACCTAG	GGTGGAGGCC	25060
	TTTCAGCCTT	CCTGCAGCAC	ATGGGGCCGA	CTGTGCACCC	TGACTGCCCG	GGCTCCTATT	CCCAAGGAGG	25130
	GTCCCACTGG	ATTCCAGTTT	CCGTACAGAG	AGGAACCCGA	ACGGCTCAGC	CACCAAGGCC	CGGTGCCCTT	25200
	CACCCCACTG	CTGAGCCAGG	GGTCTCCTGT	CCTGAGGCTC	AGAGAGGGGA	CACAGCCCGC	CCTGCCCTTG	25270
60	GGGTCTGGAG	TGGTGGGGGT	CAGAGAGAGA	GTGGGGGACA	CCGCCAGGCC	AGGCCCTGAG	GGCAGAGGTG	25340
	ATGCTGAGT	TTCTGCGTGG	CCACTGTGAG	TCTCTCGGCC	TCCACTCACA	CAGGTGGATG	TGACGGGGCG	25410
	GTACGACACC	ATCCCCAGG	ACAGGCTCAC	GGAGTTCATC	GCCAGCATCA	TCAAACCCCA	GAACACGTAC	25480
	TGCGTGCGTC	GGTATGCCGT	GGTCCAGAAG	GCCGCCCATG	GGCACGTCCG	CAAGGCCCTT	AAGAGCCACG	25550
	TAAGGTTTAC	GTGTGATAGT	CGTGTCCAGG	ATGTGTGTCT	CTGGGATATG	AATGTGTCTA	GAATGCAGTC	25620
65	GTGTCTGTGA	TGCGTTTCTG	TGGTGGAGGT	ACTTCCATGA	TTTACACATC	TGTGATATGC	GTGTGGGACA	25690
	CGTGTGTGTC	GTGGTGCATG	TATCTGTGGC	GTGCATATTT	GTGGTGTGTG	TGTGTGTGGC	ACGTGTGTGT	25760
	CCATGGTGTG	TGTGCCCTGTG	GTGTGCATGT	GTGTGTGTCT	GTGACACGTG	CATGTTTCATG	CTGTGTGCTG	25830
	CATGTCTGTG	ATGTGCCATAT	TTGTGGTGTG	TGTGTGCATG	TGTCGGTGAC	ATATGCGTGT	CTATGGCATG	25900
	GGTGTGTGTG	GCCCTTGGC	CTTACTCCTT	CCTCCTCCAG	GCATGGTCCG	CACCATTTGT	CTCACGCTCT	25970
	CGGGTGCTGG	TTTGGGGAGC	TCCACATTCA	GGGTCTCTAC	TTCTAGCATG	GGTGCCCTGT	TCTGTCTACA	26040
70	GGGCTGGGCC	TTGGAGACTG	TAAGCCAGGT	TGAGAGGAG	AGTAGGGATG	CTGGTGGTAC	CTTCTGGGAC	26110
	CCCTGGCAC	CCCAGGACCC	CAGTCTGGCC	TATGCCGGCT	CCATGAGATA	TAGGAAGGCT	GATTACAGCC	26180
	TGCTCTCCCG	GGACACACTC	CTCCAGAGC	GGCCGGGGGC	CTTGGGGCTC	GGCAGGGGTT	AAAGGGGGCC	26250
	TGGGCTTGGG	TTCCACCCCA	GTGGTCATGA	GCACGCTGGA	GGGGTAAGCC	CTCAAAGTCG	TGCCAGGCCG	26320
	GGGTGACAG	GTGAAGAAAT	ATCCCTGGAG	CTTCGGTCTG	GGGAGAGGCA	CATGTGGAAA	CCCACAAGGA	26390
75	CCTCTTTCTC	TGACTTCTTG	AGCT					26414



## Contig 2:

	TGTGGGATTG	GTTTTCATGT	GTGGGATAGG	TGGGGATCTG	TGGGATTGGT	TTTTATGAGT	GGGGTAACAC	70
	AGAGTTCAAG	GCGAGCTTTC	TTCTCTAGT	GGGTCTGCAG	GTGCTCCAAC	AGCTTTATTG	AGGAGACCAT	140
5	ATCTTCTCTT	GAACATATGGT	CGGGTTTATA	GTAAGTCAGG	GGTGTGGAGG	CCTCCCTTGG	GCTCCCTGTT	210
	CTGTTTCTTC	CACTCTGGGG	TCGTGTGGTG	CCTGCTGTGG	TGTGTGGCCG	GTGGGCAGGG	CTTCCAGGCC	280
	TCCTTGTGTT	CATTGGCCCTG	GATGTGGCCC	TGGCTACGCT	CCGTCTCTGG	AATTCCTCTG	CGAGTTGGAG	350
	GCTTCTTTTC	TTTCTTTTTT	TCITTTCTTT	TTTTTTTTTT	TGATAACAGA	GTCTCGCTCT	TTTTTGCCCA	420
	GGCTGGAGTG	GTTTGGCGTG	ATCTTGGCTC	ACTGCAACCT	GTGCTTCCTG	AGTTCAAGCA	ATTCTCTTGC	490
10	CTCAGCCTCC	CAAGTAGCTG	GAATTATAGG	CGCCACCCAC	CATGCTGACT	AATTTTTGTA	ATTTTAGTAG	560
	AGACGAGGTT	TCTCCATGTT	GGCCAGGCTG	GTCTCGAACT	CCTGACCTCA	GGTGATCCTC	CCACCTCGGC	630
	CTCCCAAAGT	GCTGGGATGA	CAGGTGTGAA	CCGCCGCGCC	CGGCCGAGAC	TCGCTTCCTG	CAGCTTCCTG	700
	GAGATCTGCA	GCGATAGCTG	CCTGCAGCCT	TGGTGTGAC	AACCTCCGTT	TTCTTCTCTC	AGGTCTCGCT	770
	AGGGGCTCTT	CCATTTCATG	ACTCTCTTCA	CAGAAGAGTT	TCACGTGTGC	TGATTTCCCG	GCTGTTTCTT	840
15	GCCTAATTGG	TGTCTGCTGT	TTATCGATGG	CCTCCTTCCA	TTTCTTTTAG	GCTTTGTTTA	TGTTGTTTTT	910
	TCCGGCTCCT	TGAAGGAAAA	GTTTCGATTA	TGGATGTTTG	AACTTTCTTT	TCTAAACAAG	CATCTGAAGT	980
	TGCCGTTTTT	CCTCTAAAGC	AGGGATCCCG	AGGCCCTTGG	CTGTGGAGTG	GCACCGGTCT	GGGGCTGTTT	1050
	AGGAACCCCG	GCGACAGCGG	GAGGCTAGGT	GGGGTGTGGG	GAGCCAGCGT	TCCCGCTTGA	GCCCGGCCCC	1120
	TCTCAGATCA	GCAAGTGGCAT	GCGGTGCTCA	GAGGCGCACA	CACCTTACTG	AGAAGTGTGC	GAGAGAGGGG	1190
20	TCTAGATTCT	GTGCTCCTTA	TGGGAATCTA	ATGCCGTGATG	ATCTGAGGTG	GAACCGTTTG	CTCCCAAAAC	1260
	CATCCCCCTC	CCCACTGCTG	TCCTGTGGAA	AAATCGTCTT	CCACGAAACC	AGTCCCTGGT	ACCACAATGG	1330
	TTGGGGACCC	TGTGCTAAAG	ACCTGCTTCA	GCAGCCTCTC	GTCACTGTTG	ATATATTTGG	TTTTCTGTGT	1400
	TGAGTCCAGA	ATAATTACGG	ATTTCTGTGA	TGCTTTCCCG	CGACCTCAGA	CCCATGGGCT	ATTTGTGGGC	1470
	GTGTTGCGTG	CTCCTGGGTT	GGGAAGGGTG	CAGGCCCCAT	GTACCTTCTT	GTTACTGCCT	TCCAGGTTGG	1540
25	TTCTCAGGGT	TGAATCGTAC	TCGATGTGGT	TTTAGCCAC	GGCCCTGCGC	CCAGCTCCTG	GGGGCTGGGG	1610
	AACATGCTGA	AGCACAGAGT	CACCGTGGCG	GTCTTTTGAT	GCCTCACAAG	CTCAGAGCCT	CCTGTGTCCG	1680
	TGTTAGTGTG	TGTACGTGTC	CTGCTCACAT	CCTGTCTTGG	GGACGCAAGG	GCTTAGCAGG	TCCCGTAGTA	1750
	AATGACAAGC	GTCCCTGGGG	AGTCTGCAGA	ATAGGAGGTG	GGGGTGGCG	TCTCTCTCCC	CGCTCTTCAG	1820
	ACTCTTCTCC	TGCCCTGTGCT	GTGGCTGCAC	CTGCATCCCT	GCAATCCCTC	CAGCACTGGG	CTGGAGAGGG	1890
30	CCGGGAGCTC	GAGTGCCACT	TGTGCCACGT	GACTGTGGAT	GGCAGTCCGT	CACGGGGGTC	TGATGTGTGG	1960
	TGACTGTGGA	TGGCGGTTGG	TCACAGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	TGGGGTCTGA	2030
	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	TGGGGTCTGA	2100
	TGTTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	2170
35	TGTTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	2240
	TGTTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	2310
	TGTTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	2380
	TGTTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	2450
	TGTTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	2520
40	GGTCTGATGT	GTGGTGACTG	TGGATGGTGA	TCGGTCACAG	GGGTCTGATG	TGTGGTGACT	TGGGTCTGATG	2590
	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTG	GTCCCGGGGG	TCTGATGTGT	GGTGACTGTG	2660
	GATGGCGGATC	GGTCACAGGG	GTCTGATGTG	TGGTGACTGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2730
	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	TGGTGACTGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2800
	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	TGGTGACTGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2870
45	GTGTTGGTGC	TGTGGATGGC	GGTCTGGGGT	TCTGATGTGT	GCTGACTGTG	GATGGCGGTC	GTGGGTCTGTG	3150
	ATGTGTGGTG	ACTGTGGATG	GTGATCGGTC	ACAGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	3220
	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGTG	3290
	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTGCTAGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCAGTCG	3360
	GTACACAGGGG	TCTGATGTGT	GGTGACTGTG	GATGGCGGTC	GTGGGGTCTG	ATGTGTGGTG	ACTGTGGATG	3430
	CGGTCGTGGG	GGTCTGATGT	GTGGTGACTG	TGGATGGCGG	TCGTGGGGTC	TGATGTGGTG	TGACTGTGGA	3500
	TGGCGGTCTGT	GGGGTCTGAT	TGGTGACTG	TGGATGGTGA	TCGGTCACAG	GGGTCTGATG	TGTGGTAGCT	3570
	GCAGGTGGAG	TCCCAGGTGT	GTCTGTAGCT	ACTTTGGGTC	CTCGGCCCCC	CGGCCCCCGT	TTCCCAAAAC	3640
50	GAAGCTTCCC	AGGCGCTCTC	TGGGCTTCAT	CCCGCCATCG	GGCTTGGCGG	CAGGTCCACA	CGTCCGTGATC	3710
	GGAGAAACAA	AGTGCCACAG	TCTGGCCGGG	GCAGGCCACA	TTTGTGGCTC	ATGCCCTCTC	CTCTGCCCGC	3780
	AGGTCTCTAC	CTTGACAGAC	CTCCAGCCGT	ACATGCGACA	GTTCGTGGGT	CACCTGCAGG	AGACCAAGCC	3850
	GCTGAGGGAT	CGCGTCTGTA	TCGAGCAGGT	CTGGGCACTG	CCCTGCAGGG	TTGGGCACGG	ACTCCACAGA	3920
	GTGGGTCCTC	CCCTGGGCAA	TCACTGGGCT	CATGACCCGA	CAGACTGTGT	GGCCTGGGGG	CGAGTGGGGG	3990
55	GAATGAGCTG	TGATGGGGGC	ATGATGAGCT	GTGTGCCCTT	GCGAAATCTG	AGCTGGGCCA	TGCCAGGCTG	4060
	CGACAGCTGC	TGCATTACAG	CACCTGCTCA	CGTTTACTG	CGCGGCTCT	CTCCAGTTCC	GCAGTGCCCT	4130
	TGTTCAATGAT	TGCTAAATG	TCTTCTCTGC	CAGTTTGTAT	CTTGAGGCCA	AAGGAAAGGT	GTCCCTCTCC	4200
	TTTAGGAGGG	CAGGCCATGT	TTGAGCCGTG	TCCTGCCACG	CTGGCCCTTC	AGTGCTGGGT	CTGAGGCCAA	4270
	AGGAAACGTG	TCCCCCTTCT	TAGGAGGACG	GGCCGTGTTT	GAGCCACGCC	CCGCTGAGCG	GGCCTCTCAG	4340
60	TGCTGGGTCT	GTCCACGTGG	CCCTGTGGCC	CTTTGACAGT	GTGGTCTGTC	CACGTGGCCC	TGTGGCTCTT	4410
	TGACAGTGCC	TGTTAGCACT	TGCTCGGCTC	TAGGGGACAG	TCGTGTCCAC	CGCATGAGGC	TCAGAGACCT	4480
	CTGGGCGAAT	TTCTTTGGCT	CCCAGGCTGG	GGGTGGAGGT	GGCCTGGGCT	GCTGGGACCC	AGACCTGTGT	4550
	CCCCGGCAGT	GGGCAGCAAC	TCCTGGATCA	CATATGCCAT	CCGGGCGCAG	GTGGGCTGTG	TGGGTGTGAG	4620
	CCCAGCTGGA	CCCACAGGTG	GCCACAGGGA	GACGTTCTGT	GTACACACT	CTGCCTAAGC	CCATGTGTGT	4690
	CTGCAGAGAC	TCGGCCCGGC	CAGCCACAGA	TGGCCCTGCA	TTCCAGCCCA	GCCCGCACT	TCATCAACAA	4760
	CAGTGACCCC	AAAAGGGACG	GAGGGTCTTG	GCCACGTGGT	CCTGCTCTGC	TCAGCACCCA	CCGGCTCACT	4830
	CCCATGTGTC	TCCCGTCTGC	TTTCGCAAG	CTCCTCCCTG	AATGAGGCCA	GCAGTGGGCT	CTTCGACGCT	4900
	TTCTACGCT	TGATGTGCA	CCAGCCGCTG	CGCATCAGGG	GCAAGTAGT	CAGGTGGCCA	GGTGCCATTG	4970
	CCCTGCGGGT	GGCTGGGCGG	GCTGGCAGGG	CTTCTGCTCA	CCTCTCTCT	GGCCCTTCCC	CAGTGNCTT	5040

	CTGCCCGGGG	CCACCAGAGT	CTCCTTTTCT	GGCCCCCGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCGA	5110
	GGCCCCGGAA	ACATGGCTCG	GCTTGCGGCA	GCCCGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGGAAATGGC	5180
5	AAGCGGGGTG	TGGAGTTGCT	CCTGCGTGGA	GGACGAGGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGCGCC	5250
	GAGCGTTTGA	GCCTGCAGCT	TGTGAGCTCC	AAGTTACTAC	TGACGCTGGA	CACCCGGCTC	TCACACGCTT	5320
	GTATCTCTCT	CTCCCCGATC	AAAAAGGATT	TATCCGATTC	TCATTCTCTG	CCCTGTCGTG	TGACCCCGCG	5390
	GAGGGCGCGG	GCTCTTCTCT	CTGTGACTAG	ATTTCCCATC	TGGAAAGTGC	GGGGTTGACC	GTGTAGTTTG	5460
	CTCCTCTCGG	GGGGCCTGTG	GTGGCCATGG	GGCAGGCGGC	CTGGGAGAGC	TGCCGTACAC	CAGCCACTGG	5530
	GTGAGCCACA	CTCACGGTGG	TAGAGCCACA	GTGCCCTGGT	CCACATCAGC	TCCTCTGGAT	TTAAGTAAA	5600
10	ACCACACACC	TCCCGGCAGG	CATCTGGCTG	CGACCTGTGT	TGTGCCCTGG	GAGAGTGGTA	GCACGGAGGA	5670
	AATTCGTGCA	CACCTAAGGT	CATCAGCAAG	GTCACTCCGA	GTCAGGTGGA	ACGTGGAGGC	CTCTCTCTGG	5740
	GATCGTCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTTCGGA	AGCTTTTATT	TAAAAATATA	ACTATTAAAT	5810
	ATTGCATTAT	AAGTAATCAC	TAATGGTATC	AGCAATTATA	ATATTATTATA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGGA	AAAACACAAA	TTGCACATGG	CAGCAGAGTG	AATTTTGCCG	GAGGGACACG	5950
15	TGTGCACATG	TGTGTAAGCG	CCCCCAGGCG	CCACAGAATT	CGCTGACAAA	GTCACCTCCC	CAGAGAAGCC	6020
	ACCACGGGGC	TCCTTCGTGG	TCGTGAATTT	TATTAAGATG	GATCAAGTCA	CGTACCGTCC	ACGTGTGGCA	6090
	GGGCTTTGGG	GAATGTGAGG	TGATGACTGC	GTCTCATGCG	CCTGACAGAC	AGGAGGTGAC	TGTGTCTGTC	6160
	CTGTCCCTAG	GACACGGACA	GGCCCGAAGC	TCTAGTCCCC	ATCGTGGTCC	AGTTTGGCCT	CTGAATAAAA	6230
	ACGTCTTCAA	AACCTGTTGC	CCCAAAACT	AAGAACAGAG	AGAGTTTCCC	ATCCCATGTG	CTCACAGGGG	6300
20	CGTATCTGCT	TGCGTTGACT	CGCTGGGCTG	GGCGGACTCC	TAGAGTTGGT	GGGTGTGCTT	CTGTGCAAAA	6370
	AGTGCACTCC	TCTTGCCCAT	CACTGTGATA	TCTGCACCAG	CAAGGAAAGC	CTCTTTTCTT	TTCTTTCTTT	6440
	TTTTTTTTTT	GAGACGGAAC	GTCACCTGTT	TCTGCCTGGG	CTTGAGTGCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GCCTCCCGGG	TTCCAGCATT	TCTCTGCCT	CAGCCTCCCC	AGCAGCTGAG	ATTACAGGCA	6580
25	CCCACCCCTC	GGCCTCTGGT	AATTTTGTGA	TTTTTAGTAG	AGAGGGGTTC	TTGCCATGTT	GGCCAGGCTG	6650
	GTCTCGAACT	CCTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAAGT	GCTGGGATTA	CAGGTGTGAG	6720
	CCATCACGCC	CAGCCGGAAA	GCCTCTTTTT	AAGGTGACCA	CCTATAGCGC	TCCCGAAAA	TAACAGGGCT	6790
	TGTTTTTGCA	GTAGGCTGCA	AGCGTCTCTT	AGCAACAGGA	GTGGCGTCTT	GTGGGCTCTG	GGGATGGCTG	6860
	AGGGTCGCGT	GGCAGCCATG	CCTTCTGTGT	GCACCTTTAG	GTTCACCGGG	GCTATTCTGC	TCTACTGTT	6930
30	TGTCTGAAAA	CGCACCCCTG	GCATCCCTTG	TGGAGAGTTC	TCTGCTCTC	GTTGGTCATG	CTGAAACTAG	7000
	GGGCAAGGTT	GTATCCGTTG	GGCGCGAGCG	GCTACATGTA	GGGTCTAGAG	TCTTTCACCG	TGGACAAATT	7070
	CCTTGAAAAA	AAAAAAAGGA	TTCGGTTAA	GCATTTCATC	CGGGTCAAGT	GTCTGGTTCT	GTGAATAAAC	7140
	CTTAAGATT	AAGAAACCTT	AATGAAAGAA	AACCTTGATG	ATTACAGAGCA	AGGATGTGGT	CACACCTGTG	7210
	GCTGGATCTG	TTTCAGCCGC	CCCACTGCTT	GGTGAGAGTG	GGGAGCAGGG	ATTGTTTGT	CAGAGGTCTG	7280
35	ATCTGGTATG	TTTCTGAGGT	GTTCGCCGCG	TGAATGGTAG	ACGTGTCGTT	TGTGTGTATG	AGGTTCTGTG	7350
	TCTGTGTGTG	GCTCGGTTTG	AGTGATCGCA	TGTCCAGCAC	ATGCCCTGCC	CGCTCTCTAC	CTGTGTCTTC	7420
	CCGCCCCCAG	TCCTACGTCC	AGTGCCAGGG	GATCCCGCAG	GGCTCCATCC	TCTCCACGCT	CTCTGCGACG	7490
	CTGTGCTACG	GGCAGATGGA	GAACAAGCTG	TTTGGCGGGA	TTCCGGCGGA	CGGGTGAGGC	CTCCTCTTCC	7560
	CCAGGGGGGG	TTGGGTGGGG	GTGATTTCG	TTTTGATGCA	TTCAAGTTTA	ATATTCTGTT	TGCTGTGAGG	7630
40	ACCATGACTG	CTCTGTCTTG	AGGAACAGCA	CAAGGTTGCA	GCCCTCTCTT	GGTATGAAGC	CGCAGGGGAG	7700
	GGGTTGACAC	GCCTGAGGAC	TGCGGGCTCC	ACGCGAGGCTC	TGTCAGCGG	CCATGTCCAG	AGAGGCTCAG	7770
	GCTCAGCAGG	CGGAGGCGCC	GCTGCCCTGC	ATGATGAGCA	TGTGAATTCA	ACACCGAGGA	AGCACACCAG	7840
	CTTCTGTGAC	GTCACCCAGG	TTCCGTTAGG	GTCCCTGGGG	AGATGGGGGT	GGTGAGCGCT	GAGGCCCCAC	7910
	ATCTCCACGC	AGGCCCTCGA	CAGGTGGGCT	GGACTGGGCG	CCTCTTCAGC	CCATTGCCCC	TCCCACTTGC	7980
45	ATGGGGTCTA	CACCCAAAGGA	CGCACACACC	TAAATATCGT	GCCAACTTAA	TGTGGTTCAA	CTCAGCTGGC	8050
	TTTTATTGAC	AGCACTTACT	TTTTTTTTTT	TAATACTTTA	AGTTCTAGGG	TACATGTGCA	CGACGTGAGC	8120
	GTATAGTTACA	TATGTATACA	TGTGCCATGT	TGGTGTGCTG	CACCCATTAA	CTCATCATTT	ACATTAGGTA	8190
	TATCTCCTAA	TGCTATCCCT	CCCCACTCCC	CCCATCCCAT	GACAGGCCCT	GGTGTGTGAT	GTTCCTCCAC	8260
	CTGTGTCCAA	GTGTTCTCAT	TGTTCACTGT	CCACCTGTGA	GTGAGAACAT	GTGGTGTGTT	GTITTTCTTC	8330
50	TTTGCAATAG	TTTGCTCAGA	GTGATGGTTT	CCAGCTTCGT	CCATGTCCCT	ACAAAGGACA	TGAACCTCAT	8400
	CTTTCTTATG	ACTGCATAGT	ATTCGGTGGT	GTATATGTGC	CACATTTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTTGG	GTGGTTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCCAAT	AAACATACGT	GTGCATGTGT	8540
	CTTTATAGCA	GCATGATTTA	TAATCCTTTG	GGTATATACC	CAGTAATGGG	ATGGCTGGGT	CAAAATGGTAT	8610
	TTCTAGTTCT	AGATCCTTGA	GGAATCACCA	CACGTCTTTC	CACAATGGTT	GAACAGTATT	ACACTCCAC	8680
55	CAACAGTGTA	AAAGTGTCTT	GGTGCTGGAG	AGGATGTGGA	CAGCAGTTAT	TTTTTATGTA	AAATAGTATC	8750
	ACTGAACAAG	CAGACAGTTA	GTGAAGGATG	CGTCAGGAAG	CCTGCAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
	GACTCCGGGT	TTTTCTGTG	CATCTTTTGA	AACCTAGACT	CCAATTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTTCAGATT	GAAATAAGTT	TATGTAACAG	AAACAAAAAT	TTCTGTGACA	8960
	GGACATTTGG	GTGGTTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCCAAT	AAACATACGT	GTGCATGTGT	9030
60	GGATACCTCT	GGCCCATGGT	CATGGGGGCG	TGGGCTTGGG	CCTGAGGGTC	ACACAGTGCA	CCATGCCACG	9100
	CTTCTCTGTG	ATAGGATCTG	GGTCTCGGAT	CATGCTGAGG	ACCACAGCTG	CCATGCTGGT	AAAGGGCACC	9170
	ACGTGGCTCA	GAGGGGGCGA	GGTTCGCCAG	CCCAGCTTTC	TTACCGTCTT	CAGTTATTTT	TCCCTAAGAG	9240
	TCTGAGAAAT	GGGGCCGCGC	CTGATGGCCT	TCGTTCTGCT	TCAGCTGGCA	CAGAAATTGCA	CAAGCTGATG	9310
	GTAACACTGT	AGTACTTATA	ATGAATGAGG	AATTGCTGTA	GCAGTTAACT	GTAGAGAGCT	CGTCTGTTGG	9380
65	AAAGAAATTT	AAGTTTTTCA	TTTAACCGCT	TTGGAGAAATG	TTACTTTTAT	TATGGCTGTG	TAAATTGTTT	9450
	GACATTACGT	CCCTCGTAGA	CAGATACTAC	GTAAGAAATG	TAAAGTTAAC	CTTGCTGTGT	ATTTTCCCTT	9520
	ATTTTAGGCT	GCTCTCGCTG	TTGTGGATG	ATTCTTGTG	GGTGACACCT	CACCTCACCC	ACCGGAAAC	9590
	CTTCTCTCAG	TGAGGCCCGT	GGCGTGTGTC	TGTGGGGACC	TCCACAGCCT	GTGGGCTTTG	CAGTTGAGCC	9660
	CCCCGTGTCC	TGCCCTTGGC	ACCGCAGCGT	TGTCTCTGCC	AAGTCTCTCT	TCTCTGCCGG	TGCTGGATCC	9730
70	GCAAGAGCAG	AGGCGCTTGG	CCGTGCACCC	AGGCTTGGGG	GCGCAGGGGC	ACCTTCGGGA	GGGAGTGGGT	9800
	ACCGTGCAGG	CCCTGGTCTT	GCAGAGAGCG	ACCCAGGTTA	CACACGTGGT	GAGTGCAGGC	GGTGACCTGG	9870
	CTCCTGCTGC	TCTTTGGAAA	GTCAAGAGTG	GCGGCTCCTG	GGGCCCCAGT	GAGACCCCCA	GGAGCTGTGC	9940
	ACAGGGCCTG	CAGGGCCGAG	GGCGCAGCCT	CCTCCCCAGG	GTGCACCTGA	GGCTGCGGAG	AGCAGGAGCT	10010
	CTGTAGGTGAG	CTGGCCACAC	CGGTCGCTG	CGGTACAGTT	CCTGCGTGGG	GTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TTGCTGAGTG	CTGCTGTCTT	GAACACCGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
75	TTTGTGAATC	AAACTAAAAT	CAGGCGACAGG	GGACCTGGGC	TCAGCACAGG	GGATTGTCCA	ATGTGGTCCC	10220
	CCTCAAGGGC	GCCCCACAGA	GCCCGTGGGC	TGTGTTTAAA	GTGCGATTGG	ACGAGGGGAC	AGAAACCTTG	10290
	AAAGCTGTAA	AGGGAACCCCT	CAGAAAATGT	GGCCGCCCAGG	GGTGGTTTCA	GGTGCTTTGC	TGGGCTGTGT	10360
	TTGTGAAAAAC	CCATTTGGAC	CCGCCCTCCA	AGTCCACCCCT	CCAGGTCCAC	CCTCCAGGGC	CGCCCTGGGC	10430

	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCCGAGC	ACAGCAGGCT	GTGCACATTT	AAATCCACTA	10500
	AGATTCACTC	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACGCTG	CTTCTGTGTG	GCAAGTTTCT	GAGGGTGCTG	GCCAGGGAGG	TGGCTCAGAG	10640
	TGTATGTGGG	GGTCCCACCG	GGGGCAGAAC	TCTGTCTCTG	ATGAGTCGGC	AGCCATGTAA	CAGGAAGGGG	10710
5	TGGCCACAGG	GAGCTGGGAA	TGCACCAGGG	GAGCTGCCCA	GCTGGCCGAG	GTCCCAGGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGCCCG	GGGCCACAGC	AGAGGCCCGA	GGAGGGGAAG	GGGATGCCCA	GGCCAGAGCA	10850
	GAGGCTACCG	GGCACAGGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCGA	GGGAACCTCC	10920
	TTGACGTGAA	GCTGACGACT	GGTGTGGCCC	AGCTCACAGC	CCAGCCAGGT	CCCGCGCCTG	AGCAGGAACCT	10990
	CAGAACCCTC	CCCTTTGTCT	AAAGCACAGC	AGATGCCTTC	AGGGCATCTA	GGAGAAAACA	GGCAAAAGTCG	11060
10	TTGAGAAACG	TCTTAAAGA	AGGTGGGATG	GTGGCAATTT	CTTGTCCAGA	TTTTAGTCTG	CCCCGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGT	GCCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCTGGGCT	GTCCCGGGTC	CAGGCCAGGT	TCTTGCATGC	TCACCTACCT	11270
	GTCTTGGCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGGA	GCAGGGCTGG	GTCCAGGCTC	CTCAGAGCTC	11340
	CTGGCAGGCC	CAGCACCCCTG	CTCCAAATCA	CCACTTCTCT	GGGGTTTTCC	AAAGCATTTA	ACAAGGGTGT	11410
15	CAGGTTACCT	CCTGGGTGAC	GGCCCCGCAT	CCTGGGGCTG	ACATTGCCCC	TCTGCCTTAG	GACCTGTGTC	11480
	CGAGGTGTCC	CTGAGTATGG	CTGCGTGGTG	AACTTGGCGA	AGACAGTGGT	GAACCTCCCT	GTAGAAGACG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCCGGCCCA	CGGCTTATTC	CCCTGGTGGC	GGCTGCTGCT	11620
	GGATACCCGG	ACCCTGGAGG	TGCAGAGCGA	CTACTCCAGG	TGAGCCGACC	TGGCCGGGAG	TGGAGCCTGT	11690
	GGCCGGCTGG	GGCAGGTGCT	GCTGCAGGGC	CGTTCGCTCC	ACCTCTGGCT	CCGTGTGGGG	CAGGCGACTG	11760
20	CCAATCCCAA	AGGCTCAGAG	GCCACAGGGT	GCCTCTCGTC	CCATCTGGGG	CTGAGCAGAA	ATGCATCTTT	11830
	CTGTGGGAGT	GAGGGTGCTC	ACAACGGGAG	CAGTTTTCTG	TGCTATTTTG	GTAAAAGGAA	GGGTGTGACC	11900
	AGACTGGGT	GCCTGAGGT	GTCTTCAGAA	AGCAGTCTGG	ATCCGAACCC	AGACGCGCCG	AGCCCTGCTG	11970
	GGCGTGAGTC	TCTCAAAACC	GAACACAGGG	GCCTTGCTGG	GCATGAGTCC	CTCTGAACCC	GAGACCTGCG	12040
	GGCCCTGCTG	GGCGTGAGTC	TCTCCGAACC	CAGAGACTTC	AGGGCCCTTT	TGGGCGTGAG	TCTCTCCGCT	12110
25	GTGAGCCCCA	CACCTCCAAGG	CTCATCCACA	GCTACAGGA	TGCCATGAGT	TCATGATCAC	GTGTGACCCA	12180
	TGAGGGGACA	GGGCCATGGT	GTGGGGGGGG	TCTCTACAAA	ATTTCTGGGG	CTTGTTCCTC	CAGAGCCCCA	12250
	GAGCTCAAAG	CCCCGTCTCA	GGCTCAGACA	CAAAATGAAT	GAAGATGGAC	ACAGATGCAG	AAATCTGTGC	12320
	TGTTTCTTTT	ATGAATAAAA	AGTATCAACA	TTCCAGGCGA	GGAAGGTGGG	CTCACACCTA	TAATCCCAGC	12390
30	ACTTTGGGAG	GCCGAGGTGG	GTGGATCACT	TGAGGCCAGG	AGTTTGAGGC	CAACCTAACC	AACATAGTGA	12460
	AATTCCATTT	CTACTTAAAA	AATACAAAAA	TTAGCCTGGC	CTGGTGGCAC	ACGCCTGTAG	TCCCGCTAT	12530
	CGGGGAGGCT	GAGGCAGGAG	AATCATTGGA	ACCCAGGAGG	CAGAGGTTGC	AGTGAGCCGA	GATCACACCA	12600
	CTGCATCCCA	GCTTGGGCAA	CAGAGTGAGA	CTTCATCTTA	AAAAAAGTATC	AAAAAGTATC	GATGTACCCA	12670
	AACCATAGTG	GACAGGTGTT	TTTTTATTCT	GTCCCTTCAT	AATATTACT	GGTGCTGTGC	TAGAGGCCCG	12740
35	AACTGGGGGT	GCCTTCTCTC	GAAGGGCACA	CCTTCATGGG	AAGAGAAATA	AGTGGTGAAT	GGTTGTTAAA	12810
	CCAGAGGTTT	AAACTGGGGT	CCTGTCTGTC	TGAGTTAACA	GTCCAGATCT	GGACTTTGCC	TCTTCCAGA	12880
	ATGCTCCCTG	GGGTTTGCTT	CATGGGGGAG	CAGCAGGTGT	GSACACCTCC	GTGATGGGGG	AGCAGCAGGT	12950
	GCAGACGCCC	TCATGATGGG	GGAGTGGCAG	GTGCAGACAC	CCTTGTGCAT	GGTGCCGAGC	ATGTCCTGT	13020
	TGCAGCTCCC	TCCCCACAAG	GATGCCGGTC	TGCTGTGCTC	CCCAAGTCC	CTGTTTCCCT	CTCACAGGCT	13090
40	TACCTGGTCC	TGGCCTCCAC	TGGCTTTGTC	TGCATGATTT	CCACATTTCC	TGGGCTCCCA	GCACCTCTTC	13160
	GCCTCTCCCA	GGCACCTCTG	CAGTGCTGGC	CATACCACTG	AGCTGTGAAC	TGTCCACTGC	TTATTTTGTCT	13230
	CCCCATGAAA	TGTATTTTTT	AGGACAGGCA	CCCTGGTTTC	CAGCCTCTGG	CACAGCATCA	GTGAATCTTA	13300
	TTGAAGGACA	AAGGACAGAC	AAACAATCA	GGAAAATGGG	TTCTCTCTAA	ACACATTGCA	AAGCCACAGA	13370
	GGCTAGTGCA	GGATGGGTGG	GCATCAGGTC	ATCAGATGTG	GGTCCAAATG	CAGAAATATC	TGTGCTCCCA	13440
45	AAGGCCACTT	GGTCAGAGTG	TGTGCTTGCA	GAGGTGGCTC	TAAAAGCTCA	GCAGTGGAGG	CAGTGGTTCG	13510
	CCATACTCAG	GGTGAATCTA	CATCTCTGT	GTCTGAAATA	TACAGCAGAG	GCTTGAAGGG	CTCTGGGAG	13580
	AAGAAAACAG	GCAAAATGAT	TAAGAAAAGT	GAAAAGGAA	AAGTGGTAAG	ATGGGAATTT	TCTTGTCCAG	13650
	ATTTTAGTCT	CCCAAAACCAC	AGCTCAGATG	GTAGAATGTG	GTCAGAACTG	ATGGACAGAA	CAATAGAACA	13720
	AAACGGGAAG	CCTATCTCTC	AGAAACGTGT	GTAAATGTGG	TATGTGGCAC	AGCTGATGGA	AAAGAGAGTG	13790
50	TGTGTGTAAT	TTTTTTTTCT	GAGAAAACCTG	ACTGGAAGCA	AATAAGTTGT	GTCTTTACAG	CATATACCAG	13860
	AGCAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAAAACA	CCAGCAACAG	AAATAAAACA	AAAGACTCAA	13930
	AGGGGAAGGA	GGTGAACGTT	CCCTGGTTTG	GTGTTGGGGA	AGGACACACA	GGGAGGGCGA	TGAAACCACT	14000
	GAGGCAACGG	GCATTGCTTT	CACCTGCAGAG	AAACTCAGCT	TGCCTGAGCC	ACAGTGAAAA	TGGCCATTCC	14070
	CTGGAGCGTT	TGTGCACGTG	ATTTATTTAA	GGCGCCCTGT	GAGGTCTCTG	ACATTCTATC	TCTCACTTTG	14140
55	TTCTCTTAAC	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCCA	GGGAGGACGC	CGCCCTTGGT	CACCCAGCTG	14210
	GCAAAAGGCA	TGCATGATTG	CAGCCTGGCC	TCTGTCTCCG	GGGCCCTTGC	TCTGCCCGAG	GACCCACAC	14280
	AAGTCAGACC	CATAGGCTCA	GGGTGAGCCG	GAGCCCAAGG	TCGTGTTGGG	GATGGCTGTG	AAAGAAGAAA	14350
	TGGAGCTCTG	ATGCACACTT	GGGAAGGTCC	TACCAGCAGC	GTCAAAGAAA	TGCATGTGAA	ACTGACAGCG	14420
	AGACCCATCC	CTCAAAGAAA	CGCACGTGAA	ACTGATGGCG	AGACCTGTCC	CCATCCCTCA	TGCTGGCTCC	14490
60	TTTTCTGGGC	TTGCCAAGAG	CCAGCATCAG	GTTGAGGCAA	GCTGGAAAGA	CTTTTCTGGA	AAGCAGCTTG	14560
	TTTGCATGGA	AGTCCTCACA	ATGTCTCTGT	TCTTCCCACT	AATTCCACTT	CTGAAGTGAC	CAGACATTAT	14630
	CAGGGGTCTT	ATTTACCATT	TCCAGTGTTC	CAGGCAAGGG	GACTTGCCAC	AGCAAGTCAC	GAACTGTCCC	14700
	AAATACAGGG	CTAAGGAGAT	ATTATGCATC	ACAAAACCTT	CTCTGCCATT	AAACATTTTT	CAAGAATTTT	14770
	TTGAAGAATG	TTTAATGGCA	CAAAACGTTT	ATTTCAATGT	AGCAGTGTTC	AAAGCTGGAT	GTAAAAGAAC	14840
65	ACACCCAGGG	AGCCTGCCGT	GAATGTCATG	TGTGTTTATC	TTTGACATG	GACATACATG	GGCAGTGAGT	14910
	GGTGGTGAGG	CCCTGGAGGA	CATCGGTGGG	ATGCCCTCCAT	CCTGCCCCCT	TGGAGACACC	ATGTGTGCCA	14980
	CGTGCACTCA	CTGGAGCCCT	GTTTAGCTGG	TGCCACCTGG	CTCTTCCATC	CCTGAGATTC	AAACACAGTG	15050
	AGATTCCCCA	CGCCCAACTC	AGTGTCTCTC	CACAAAAAAC	CTGAGTCACA	CCTGTGTTCA	CTCGAGGGAC	15120
	GCCCGGGAGC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTTTG	GCTGAGTTAT	GTGCAGATCT	CATCAGGGCA	15190
70	GATGATGAGT	GCACAAACAC	GGCCGTGCCA	GGTTGGGATA	CACCTCAACAT	CACCTAGCCAG	GTCCGTGGTG	15260
	AGTTTGGTCA	TGCAGAGTCT	GGATGGCATG	TAGCATTGGG	AGTCCATGGA	GTGAGCACCC	AGCCCCCTCG	15330
	GGCTGCAGCG	CATGCCCCAG	GCAGGACAAG	GAAGCGGGAG	GAAGGCAGGA	GGCTCTTTGG	AGCAAGCTTT	15400
	GCAGGAGGGG	GCTGGGTGTC	GGGCAGGCAC	CTGTGCTGTA	CATTCCCCCT	TGTGTCTGAG	CTATGCCCGG	15470
	ACCTCCATCA	GAGCCAGTCT	CACCTTCAAC	CGCGGCTTCA	AGGCTGGGAG	GAACATGCGT	CGCAAACTCT	15540
75	TTGGGGTCTT	GCGGCTGAAG	TGTTCACAGC	TGTTTCTGGA	TTTGCAGGTG	AGCAGGCTGA	TGGTCAGCAC	15610
	AGAGTTTCAGA	GTTTCAGGAG	TGTGTGCCCA	AGTATGTGTG	TGTGTGTGTG	CGCGCGTGCC	TGCAAGGCTG	15680
	ATGGTGACTG	GCTGCACGTA	AGAGTGCACA	TGTACGCATA	TACAGCTGAG	CACATACATG	TGTGCATGTG	15750
	TGTACATGAA	GGCATGGCAG	TGTGTGCACA	GGTGTGCAAG	GGCACAAAGT	TGTGCACATG	CGAATGCACA	15820

	CTGACATGC	ATGTGTGTTT	GTGCACAGTC	GTGTGGGCAT	TCACGTGAGG	TGCATGCGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTC	ACCCCGCTAG	GTCTCAGCA	15960
	CCAGTGCCAC	TCCTTACAGG	ATGAGACGGG	GTCCCAGGCC	TTGGTGGGCT	GAGGCTCTGA	AGCTGCAGCC	16030
5	CTGAGGGCAT	TGTCCTATCT	GGGCATCCGC	GTCCACTCCC	TCTCCTGTGG	GCTTCTGTGT	CCACTCCCCC	16100
	TCTCCTGTGG	GCATTTACAT	CCACTCCACT	CCCTCTCTCC	TGTGGGCATC	CGCGTCCACT	CCCCCTCTCT	16170
	TGGGGCATCT	CGCTCCACCT	CCCCCTCTCT	TGGGCATTGG	CGTCCACTCC	CTCTCCTGGT	TCCTTCTCTGT	16240
	CTTGGCCGAG	CCTCGGGGGC	AGGCAGATGA	CACAGAGTCT	TGACTCGCCC	AGGGTGGTTC	GCAGCTGGCC	16310
	GGTGAGGGCC	AGGCCGGATT	TCACTGGGAA	GAGGGATAGT	TTCTTGTCAA	AATGTTCTCT	TTTCTTGTTC	16380
10	CATCTGAATG	GATGATAAAG	CAAAAAGTAA	AAACTTAAAA	TCCCAGAGAG	GTTTCTACCG	TTTCTCACTC	16450
	TTTCTTGGCG	ACTCTAGGTG	AACAGCTCC	AGACGGTGTG	CACCAACATC	TACAAGATCC	TCCTGCTGCA	16520
	GGCGTACAGG	TGAGCCGCCA	CCAAGGGGTG	CAGGCCAGC	CTCCAGGGAC	CCTCCGCGCT	CTGCTCACTT	16590
	CTGAGCCCGG	GCTTACCTTT	GGAACCTCTG	GGTTTTAGGG	GCAAGGAATG	TCTTACGTTT	TCAGTGGTGC	16660
	TGCTGCCTGT	GCACAGTTCT	GTTCGCGTGG	CTCTGTGCAA	AGCACCTGTT	CTCCATCTCT	GGGTAGTGGT	16730
15	AGGAGCCGGT	GTGCCCCCAG	GTGTCCCCAC	TGTGCTGTG	CACCTGGCCGT	GGGACGTCAT	GGAGGCCATC	16800
	CCAGGGCAGC	AGGGGCATGG	GGAAGAGAGA	TGTTTATGGG	GAGTCTTAGC	AGAGGAGGCT	GGGAAGGTGT	16870
	CTGAACAGTA	GATGGGAGAT	CAGATGCCCG	GAGGATTTGG	GGTCTCAGCA	AAGAGGGCCG	AGGTGGGTGC	16940
	AGGTGAGGGT	CGCTGGCCCC	ACCCCGGGGA	AGGTGCAGCA	GAGCTGTGGC	TCCCAGACAC	GGCCGGCCAG	17010
	CACCTGTGCT	CTGGGCATGG	CTGTGCTCCT	GGAACGTTCC	CTGTCTGTGG	TGGTCAGGGG	GTGCCCTGCT	17080
20	CAAGAATCGA	CAACTTTATC	ACAGAGGGAA	GGGCCAATCT	GTGGAGGCCA	CAGGGCCAGC	TTCTGCCTGG	17150
	AGTCAGGGCA	GGTGGTGGCA	CAAGCCTCGG	GGCTGTACCA	AAAGGCGATC	GGGCACACAC	GGCCCGGGCC	17220
	TCCACCTCAA	CAGGCCTCCC	GAGCCACTGG	GAGCTGAATG	CCAGGAGGCC	GAAGCCCTCG	CCCCATGAGG	17290
	GCTGAGAAAG	AGTGTGAGCA	TTTGTGTAC	CCAGGGCCGA	GGCTGCGCGA	ATTACCTGTC	ACACTTGATG	17360
	TGAATGAGG	TCGTGCTCTA	CGTGGAAAC	CCAGCAAGGG	CTCACGGGAG	AGTTTTCCTAT	TACAAGGTGC	17430
25	TACCATGAAA	ATGGTTTTTA	ACCCGAGTGC	TTGCGCCTTC	ATGCTCTGGC	AGGGAGGGCA	GAGCCACAGC	17500
	TGCATGTTAC	CGCCTTTTGA	CCAGCTCCAG	AGGCTTGGGA	CCAGGCTGTC	TCAGTTCCAG	GGTGGCTCCG	17570
	GGTCAGACCG	CCCTCCTCTC	TGCTTCTCTC	CTCTGCCTCA	AATCTTCCCT	CGTTTGCTATC	TCCTTGAGCG	17640
	GTGCTGTGGC	CCTCGTGCAA	CGTCTTGAC	TCCTTTCCGG	AAACCCCTTG	GGTGTGCTGG	ATACAGGTGC	17710
	CACAGGAGAC	TGGAGGTGTC	TGACACTGTG	GTTGACCCCA	GGGTCCAGCT	GGCGTGCTTG	GGGCTCCTTT	17780
30	GGGCATGAT	GAGGTGAGAG	GAGTTTTCCC	AGGTGAAAC	TCCTGGGAAA	CTCCCAGGGC	CATGTGACTG	17850
	GGCACCCTGT	CTCCCATAT	TCAGCTCAGT	CTTGTCCCTA	TTTCCCCACC	AGGGTCTCTA	GCTCCGAGGA	17920
	GCTCCCGTAG	AGGGCCCTGG	CTCAGGGCAG	GGCGGCTGAG	TTTCCCCACC	CATGTGGGGA	CCCTTGGGTA	17990
	GTGCTGTGAT	TGGGTAGCCC	TGAGGAGGCC	GAGATGCCAT	GGGCCACGGG	CCGTTTCCAA	ACACAGAGTC	18060
	AGGCACGTGG	AAGGCCCAGG	AATCCCTTTC	CCTCGAGGCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCGGA	18130
35	TTTACCGGCA	GCCAGGCTGC	AGTGGCGGAG	GCTGTGGTGG	TCCACGTGGC	GCTGGGGGGC	GGGTCTGATT	18200
	CAAACTCCGT	GGGGCTCGGC	CTTCTCGGCC	CGTGTGGGCC	GGCGCTCCAC	ACGGGCTTGG	GTGGGACGCC	18270
	CCGACCTCTA	GCAGGTGGCT	ATTTCTCCCT	TTGGAAGAGA	GCCCTCACCC	CATGCTAGGT	GTTCCTCTGC	18340
	CGGGTGCTGT	GCTTGTGCTG	AAATGGGAAA	AAGACATCCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAAC	18410
40	TACATGGGGG	GCTCAGGCAG	TGGGTGAGAT	GAGGTACACG	GGGGCTCAG	GCAGTGGGTG	AGGCCAGGTA	18480
	CATGGGGGGC	TCAGGCACCTG	GGTGAGATGA	GGTACACGGG	GGGCTCAGGC	AGAGGGTCAG	ACCAGGTATG	18550
	CGGGGGCTCT	GATCACACGC	ACATATGAGC	ACATGTGCAC	ATGTGCTGTT	TCATGGTAGC	CAGGTCTGTC	18620
	CACACCTGCC	CCAAAGTCCC	AGGAAGCTGA	GAGGCCAAGG	ATGGAGGCTG	ACAGGGCTGG	CGCGTGGCTT	18690
	CACACCTGTA	GTCCAGCAGC	TTTGGGAGGC	CGAGGCGAGA	GGATCCCTTG	AGCCACAGGAG	TTTAAGAGCA	18760
45	CCCTGAGCAA	CATAGTAGAA	CCCCATCTCT	ATGAAAAATA	AAACAAAAAA	TAGCTGAAC	ATGGTGGTGT	18830
	GGCCTGTGAT	TTCCAATACT	TGGGAGGCTG	AAGTGGGAGG	ATCACTTGAG	CCCAGGAGGT	GGAGCTGATG	18900
	GTGAGCTGAG	ATTGCACCAC	TGTACTGCAG	CCTGGGTGAC	AGAGTGAGAG	CCCATCTCAA	CAACAACAAA	19040
	GAAGACTGAC	AAATGCAAGT	TCCTGGAAAG	AAACATTTAG	TAGGAACCTA	ACCTACACAC	AGAAGCCAGG	19110
50	TCGGTGTCTC	GGTGTCACTG	AGATGAGATG	ATGGGTCTCT	ACACCTAC	CCCAGACCCA	GGGTTTATGC	19180
	ACCACAGGGG	CGGGTGGCTC	AGAAGGGATG	CGCAGGACGT	TGATATACGA	TGACATCAAG	GTGTCTGAGC	19250
	GAAGGGCAGG	ATTATGATA	AGTACTGTCT	GGTACACAA	GAACAATGGA	TAACTGGAA	ACCTTAGAGG	19320
	CCTTCCCGGA	ACAGGGGCTA	ATCAGAAAGC	AGCATGGGGG	GCTGGCATCC	AGGATGGAGC	TGCTTACGCC	19390
	TCCACATGCG	TGTTTATACA	GATGGTGCAC	AGAAACGCAG	TGTACCTGTG	CACACACAGA	CACGCAGCTA	19460
55	CTGCGACACA	CAAGCACACA	CACAGACATG	CATGCATGCA	TCCGTGTGTG	TGCACCTGTG	CCCATGAGGA	19530
	AACCCATGCA	TGTGCATTCA	TGCACGCACA	CAGGCACCGG	TGGGCCCATG	CCCACACCCA	CGAGCACCGT	19600
	CTGATTAGGA	GGCCTTTTCT	CTGACGCTGT	CCGCCATCCT	CTCAGGTTTC	ACGCATGTGT	GCTGCAGCTC	19670
	CCATTTCATC	AGCAAGTTTG	GAAGAACCCC	ACATTTTCTC	TGCGCTGTCAT	CTCTGACACG	GCCTCCCTCT	19740
	GCTACTCCAT	CCTGAAAGCC	AAGAAGCCAG	GTATGTGCAG	GTGCTTGGCC	TCAGTGGCAG	CAGTGCCTGC	19810
60	CTGCTGGTGT	TAGTGTGTCA	GGAGACTGAG	TGAATCTGGG	CTTAGGAAGT	TCTTACCCCT	TTTCCGATCA	19880
	GGAAGTGGTT	TAACCCAAAC	ACTGTCAAGC	TCGTCTGCCC	GCCTCTCTGT	GGGGTGAGCA	GAGCACCTGA	19950
	TGGAAGGGAC	AGGAGCTGTC	TGGGAGCTGC	CATCCTTCCC	ACCTTGTCTT	GCCTGGGGAA	CGCTGGGGGG	20020
	GCCTGGTCTC	TCCTGTTTTC	CCCATGGTGG	GATTTGGGGG	GCCTGGCCCT	TCCTGTTTTC	CCTGTGGTGG	20090
	GATTGGGCTG	TCTCCCGTCC	ATGGCACTTA	GGGCCCTTGT	GCAAAACCCAG	GCCAAAGGGCT	TAGGAGGAGG	20160
65	CCAGGGCCAG	GCTACCCACG	CCCTCTCAGG	AGCAGAGGCC	GGTATACACC	ACGACAGAGC	CCCGCGCCGT	20230
	CCTTCTGCTT	CCAGTCAACG	TCCTCTGCCC	CTGGACACTT	TGTCCAGCAT	CAGGGAGGTT	TCTGATCCGT	20300
	CTGAAATCCA	AGCCATGTCT	AACCTGGCGT	CCTGAGCTTA	ACAGCTTCTA	CTTTCTGTTC	TTTCTGTGTT	20370
	TGGGAAATTT	CACCTGGAGA	AGCCGAAGAA	AACATTTCTG	TCGTGACTCC	TGCGGTGCTT	GGGTCCGGAC	20440
	AGCCAGAGAT	GGAGCCACCC	CGCAGACCGT	CGGGTGTGGG	CAGCTTTCCG	GTGTCTCCTG	GGAGGGGAGC	20510
70	TGGGCTGGGC	CTGTGACTCC	TCAGCCTCTG	TTTTCCCCCA	GGGATGTCCG	TGGGGGCCAA	GGGCGCCGCC	20580
	GGCCCTCTGC	CCTCCGAGCC	CGTGCACTGG	CTGTGCCACC	AAGCAATTCCT	GCTCAAGCTG	ACTCGACACC	20650
	GTGTACCTTA	CGTGCCACTC	CTGGGGTCA	TCAGGACAGG	CAAGTGTGGG	TGGAGGGCCG	TGCGGGCCCC	20720
	ACCTGCCCCAG	GGGTCACTCT	TGAACGCCCT	GTGTGGGGCG	AGCAGCCTCA	GATGCTGCTG	AAGTGCAGAC	20790
	GGCCCGGGGC	CTGACCCCTG	GGCCCTGGAG	CCACGCTGGC	AGCCCTATGT	GATTAAACGC	TGGTGTCCCC	20860
	AGGCCACCGA	GCCTGGCAGG	GTCCCAACT	TCTTGAACCC	CTGCTTCCCA	TCTCAGGGGC	GATGGCTGAC	20930
75	CACGCTTGGG	AGCCTTCTGA	CCCTGACCT	GTGTCTCTCT	ACAGCCTCTT	CCCTGGCTGC	TGCCCTGAGC	21000
	TCCTGGGGTC	CTGAGCAAGT	TCTTCTCCCG	CCCCGCGCGT	CCAGCGTCAC	TGGGCTGCCT	GTCTGTCTGC	21070
	CCCGGTGGAG	GGGTGTCTGT	CCCTTCACTG	AGGTTCCAC	CAGCCAGGGC	CACGAGGTGC	AGGCCCTGCC	21140

TGCCCCGCCA CCCACACGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTTCTGGA ACGGAGTCTG 21280  
 ATTTTGGCCCC CGCAGCCACG ACGCAGCTGA GTCCGAAGCT CCCGGGGACG ACGCTGACTG CCCTGGAGGC 21350  
 CGCAGCCAAC CCGGCACTGC CCTCAGACTT CAAGACCATC CTGGACTGAT GGCCACCCGC CCACAGCCAG 21420  
 GCCGAGAGCA GACACCAGCA GCCCTGTAC GCCGGGCTCT ACGTCCCAGG GAGGGAGGGG CGGCCACAC 21490  
 CCAGGCCCGC ACCGCTGGGA GTCTGAGGCC TGAGTGAGTG TTTGGCCGAG GCCTGCATGT CCGGCTGAAG 21560  
 GCTGAGTGT CCGGCTGAGG CTGAGCGAGT GTCCAGCCAA GGGCTGAGTG TCCAGCACAC CTGCCGTCTT 21630  
 CACTTCCCCA CAGGCTGGCG CTCGGCTCCA CCCCAGGGCC AGCTTTTCTT CACCAGGAGC CCGGCTTCCA 21700  
 CTCGCCACAT AGGAATAGTC CATCCCCAGA TTCGCCATTG TTCACCCCTC GCCTTGGCCTT CTTTGGCCTT 21770  
 CCACCCCCAC CATCCAGGTG GAGACCCCTG GAAGGACCCCT GGGAGCTCTG GGAATTTGGA GTGACCAAG 21840  
 GTGTGCCCTG TACACAGCGG AGGACCCCTG ACCTGGATGG GGGTCCCTGT GGGTCAAAAT GGGGGGAGGT 21910  
 GCTGTGGGAG TAAAATACTG AATATATGAG TTTTTCAGTT TTGAAAAAAA TCTCATGTTT GAATCCTAAT 21980  
 GTGCACTGCA TAGACACCAC TGTATGCAAT TACAGAAGCC TGTGAGTGAA CGGGGTGGTG GTCACTGCGG 22050  
 GCCCATGGCC TGGCTGTGCA TTTACGGAG TCTATGAGTG AATGGGGTTG TGGTCACTGC GGGCCCATGG 22120  
 CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGCAGG AGGGGAAGGA GGGTAGGGGA TAGACAGTGG 22190  
 GAGCCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGCCG GAAGGGCAGC AGGGATGCTG GGGGCCAGC 22260  
 TTGGGCGGCG GGGATGATGG AGGGCTGGC CAGGGTGGCA GGGATGATGG GGGCCCGCAG TGGGTGGCA 22330  
 GGGGTGATGG GGGGGGCTGG TCTGGGTGGC GGGGAAGATG GGGGAAGCTG GCTGGGCCCT CTCCTCCCTT 22400  
 GCCTCCACCC TGCAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTTCTT 22470  
 GGAGGTGGGG GGCAGGGGCA TGACACCATC CTGTATAAAA TCCAGGATTC CTCCTCTCTG ACGCCCCAAC 22540  
 TCAGGTGAA AGTCACATTC CGCCTCTGGC CATCTCTTAA AGAGTAGACC AGGATCTCTG TCTCTGAAGG 22610  
 GTGGGTAGGG TGGGGCAGTG GAGGGTGTGG ACACAGGAGG CTTAGGGGTG GGGCTGGTGA TGCTCTCTCA 22680  
 TCCTCTTATC ATCTCCAGT CTCTCTCTCT ATCTCTTAT CATCTCCAG TCTCATCTGT CTCTCTCTTA 22750  
 TCTCCAGTCT TCATCTGTCA TCCTCTTACC ATCTCCAGT CTCTCTCTT ATCTCTTAT TCCTTAGTCT 22820  
 CATCCAGACT TACCTCCAG GGGGGGTGCC AGGCTCCGAG TGGAGCTGGA CATACGTCCT TCCTCAGGCA 22890  
 GAAAGGAATG GAAGGATTC AGAGAACAGG AGGGGCGGCT CAGAGGGAGC CAGCTTTGGG GTGAAGAAAC 22960  
 AGCCCTCTCT CAGAAGTTGG CTTGGGCCAC ACGAAACCGA GGGCCCTGCG TGAGTGGCTC CAGAGCCTTC 23030  
 CAGCAGGTCC CTGTTGGGCG CTTATGTTAT GGCCGGGTCC TACTGAGTGC ACCTTGGACA GGGCTTCTGG 23100  
 TTTGAGTGCA GCCCGGACGT GCCTGGTGTG GGGGTGGGGG CTTATGGCCA CTGGATATGG CGTCATTTAT 23170  
 TGCTGTGCTC TCAGAGAATG TCTGAGTGAC CGAGCCTAAT GTGTATGGTG GGCCCAAGTC CACAGACTGT 23240  
 GTCGTAAATG CACTCTGGTG CTTGGAGCCC CCGTATAGGA GCTGTGAGGA AGGAGGGGCT CTTGGCAGCC 23310  
 GGCTTGGGGG CGCCTTTGCC CTGCAAACTG GAAGGGAGCG GCCCGGGGCG CCGTGGGCGG ACGACCTCAA 23380  
 GTGAGAGGTT GGACAGAAAC GGGCGGGGAC TTCCAGGAG CAGAGGCCGC TGCTCAGGCA CACCTGGGTT 23450  
 TGAATCACAG ACCAACAGGT CAGGCCATTG TTCAGCTATC CATCTCTAC AAAGCTCCAG ATTCCTGTTT 23520  
 CTCGGGTGT TTTTGTGTA AATTTTACTC AGGATTACTT ATATTTTGT CTAAAGTATT AGACCTTAA 23590  
 AAAAGGTATT TGCTTTGATA TGGCTTAACT CACTAAGCAC CTACTTTATT TGCTGTGTTT TATTTATTAT 23660  
 TATTATTATT ATTAGAGATG GTGTCTACTC TGTCACCCAG GTTGTAGTG CAGTGGACA GTCATGGCTC 23730  
 GCTGTAGCCG CAAACCCCA GGCTCAAGTG ATCTCCGGC CTCAGCTTCC CAGAGTGCTG GGATTACAG 23800  
 TGTGAGCCAC TGCCCTTGCC TGGCACTTTT AAAAACCACT ATGTAAAGTC AGGTCCAGT GCTTCCACAC 23870  
 CTGTATCCC AGTAGTTTG GAAGCCGAGG CAGAAGGATT GTCTAGGCC AGGAGTTTGA GACCAGCATG 23940  
 GGTAAACATG GGAGACCCCA TCTCTACAAA AAATGCAAAA AGTTATCCGG GCGTGGGTC CAGCATCTGT 24010  
 AGTCCCAGCT GCTCGGGAGG CTGAGTGGGA GGATCGCTTG AGCCCGGGAG GTCATGGCTG CAGTGAGCTG 24080  
 TGATGTATCC ATCGCACTCC AGCCTGGGCA ACAGAGTGAG ACCCTGTCTC AAAAAAAGG AAAAAAAG 24150  
 AAGGAGAAGG AGAAGAGAAG AAGAAGGAAG AAGGAAAGAG AAGAAGAAG AAGAAGGAAG AAGAAGGAG 24220  
 AAGGAGGCTT GCTAGGTGCT AGGTGAGCTG TCAATCTCA GAGCAAAATG AAATATACAA AGTTTAAAG 24290  
 GGAAAGAAAA ACCCAGCTC TTTGGACTTC CTTAGGCCCTG AACTTCACTC CAAGCAGCTT CTTTCCACAG 24360  
 ACAAGCGTGT ATGGAGCGAG TGAGTTCAAA CGAGAAAGGG AGGAGAAGCA GGAAGGGTG GAGGCTGTGG 24430  
 GTGACACCAG CCAGGACCCC TGAAGGGGAG TGGTTGTTTT CTTGCTCAG CCCACGCTC CTGCCGCTC 24500  
 TGCACCTGCT GTAACCGTCG ATGTTGTTGC CAGGTGCCCA CTTGGGAAGG ATGCTGTGCA GGGGGCTTGC 24570  
 CAAACTTTTG TGGGTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAAAT ACAGCCCTC CCAAGAGATG 24640  
 CCCACGTCTT TCTCTGGAA CCTGTGAATG GTGCACCCGC AAGGCAGAGG CTGGTGAAGG CTGCAGGTGG 24710  
 AATCACGGCT GCCAGTCAGC CGATCTTAAG GTCATCTGG ATTATCTGGT GGGCCTGATA TGGCCACAG 24780  
 GGTCCCTAGA AGTGAGAGAG GGAGGCAGGG GAGACTCAGA GAGGGGACGT GAGAAGGACC ACTGCCACT 24850  
 GCTGGCTTTG AGATGGAGGA GGGGGTCCCC AGCCAAGGAA TGGGGGACG CGCTCCATGC TGGAAAAGCA 24920  
 AGCAATCTCT CCCGTCCTG AGGGCACACG GCCCTGCCCA CGCCTCGATT TCAGGCCAGT GGGACCTGTT 24990  
 TCAGCTTTCC GGCTCCAGA GCTGTAAGAT GATCGGTTTG TGTTCAGCCA CTAAGCTGCA GTGATTCGTC 25060  
 ACAGCAGCAA ATGGAATAGC AGTACAGGGA AATGAATACA GGGACAGTTC TCAGAGTGAC TCTCAGCCCA 25130  
 CCCCTGGG 25138

## 60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz  
 der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der  
 Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist  
 in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in  
 5 den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon			Intron				3'-Exon			
Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N C	A G   G
10 Häufigk.(%)	70	60	80		100	100	95	70		80	100 100 60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616  
 15 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position  
21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

## Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCCGGGTCGCGCTCCGGCTGGGGTTGAGGGCGGCCGGGGGAACAGCGACATGCGGAGAGCAGCGCAGG  
CGACTCAGGGCGCTTCCCCCGCAG

5

## Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTCGAGGGCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCC  
CTGGTCCTCCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTCGAGTGGACACGGTGATCTCTGCC  
TCTGTCTCCTCTCCTGTCCAGTTTGCATAAACTTACGAGGTTACCTTTCACGTTTTGATGGACACGCGGTTTCCAGGCGC  
CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGGTTGCCGCAATGGGGAGAAAGTGTCTGGAAG  
CACAGACGCTCTGGCGAGGGTGCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGGTGGGAATGAGAGGTGGGGA  
CGAGAACCCCTCTTCTGGGGGTGGGAGGTAAGGGTTTTGCAGGTGCACGTGGTCAGCCAATATGCAGGTTTGTGTTTA  
AGATTTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGGTAATCCAGCACTTTGGGAAGCTGAGGCAGGTGGA  
TCACCTGAGGTGAGGAGTTTGAGACCAGCTGACCAACATGGTGAACCCCTATCTGTACTAAAAATACAAAATTAGCTG  
GGCATGGTGGTGTGTGCTGTATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTGAACCCAGGAGGCGGAGGC  
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAGAGTGAACTCTGTCTTAAAAAAAAGTGTT  
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCCTCTTTAGGTAT  
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGGTGGTGTTCAGGGG  
ATGGTGTCTGTGGGCCCTGCCGTGTCCCCACCCTGTTTTCTGGATTGTATGTTGAGGAACCTCCGCTCCAGCCCCCTTT  
TGGCTCCCAGTGCTCCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCGATTTCACCCCTCCCCACAACTCCCAAGAC  
ATGTAAGACTTCCGGCCATGCAGACAAGGAGGGTGACCTTCTGGGGCTCTTTTTTTCTTTTTTTCTTTTATGGTGGC  
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGCAGAATTGCTAACTCGGCGGTGTTTA  
CAGCAGGTGTCTTGAAATGCTGCGTCTTGCCTGACTGGAAGTCCCTACCCATCGAACGGCAGCTGCCTCACACCTGTCTGC  
GGCTCAGGTGGACCACGCCAGTCAGATAAGCGTCATGCAACCCAGTTTTGCTTTTTGTGCTCCAGCTTCTCTCGTTGAG  
GAGAGTTTGAGTTCTCTGATCAGGACTCTGCCTGTATTGCTGTTCTCTGACTTCAGATGAGGTACAATCTGCCCTCGG  
CTTATGCAGGGAGTGAGGCGTGGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT  
GTACAGTGTAGGGTGAGTGAGGCGCGGCCCGGGTGTCCCTGTCCCGTGCAGCGTGATTGAGGTGTGGCCCCGGGTGT  
CCCTGTACAGTGTAGGGTGAGTGAGGCGCCATCCCCGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCGTGGTCCCCGG  
GTGTCCCTGTCCCGTGCAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGCGGTCC  
CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGGCCCCAGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCACC  
GTCCCTGGGTGTCCCTCCCAGGTATAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGG  
CGCGGCCCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT  
GAGGCTCTGTCCCAGGTGTCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGCTCTTTCTATAGCCACAGCT  
GCGCCGGTTGCCATTGCCTGGGTAGATGGTGCAGGCGCAGTGTGGTCCCCAAGCCTATCTTTCTGATGCTCGGCTCT  
TCTTGGTCACCTCTCCGTTCCATTTTGCTACGGGGACACGGGACTGCAGGCTCTCGCTCCCGGTGCCAGGCACTGCAG  
CCACAGCTTCAGGTCCGCTTGCCCTCTGTGGGCTGGCTGTCTACCAAGTGCCCGCCACATGCATGCTGCCAATACTCC  
TCTCCAGCTTGTCTCATGCCAGGCTGGACTCTGGGCTGCCTGTGTCTGTGCCACGTGTTGCTGGAGACATCCAGAA  
AGGGTTCTCTGTGCCCTGAAGGAAAGCAAGTCACCCAGCCCCCTCACTTGTCTGTTTTCTCCCAAGCTGCCCTCTGC  
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGTACCTTATTCTGGGCACCTGCCGCTCATTGCTTAGGCTGGGCTCTGCCT  
CCAGTCCGCCCTCACATGGATTGACGTCCAGCCACAGGTTGGAGTGTCTCTGTCTGTCTCTGCTCTGAGACCCACGTG

40



GAGGGCCGGTGTCTCCGCCAGCCTTCGTGAGACTTCCCTCTGGGTCTTAGTTTGAATTTCACTGATTTACCTCTGACG  
TTTCTATCTCTCCATTGTATGCTTTTCTGGTTTATTCTTTTCATTCTTTCTAGCTTCTTAGTTTAGTCATGCCTTTC  
CCTCTAAGTGCTGCCTTACCTGCACCCCTGTGTTTTGATGTGAAGTAATCTCAACATCAGCCACTTTCAAGTGTCTTAAA  
5 ATACTTCAAAGTGTAAATACTTCTTTTAAGTATTCTTATTCTGTGATTTTTTCTTTGTGACGCTGTGTTTTGACGTGA  
AATCATTTTGATATCAGTGACTTTTAAGTATTCTTAGCTTATTCTGTGATTTCTTTGAGCAGTGAGTTATTTGAACACT  
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT  
TGTTAGTGGTCTGTATAATACCAATTATTTGAAGTTTGGGAGCCTTGCTTTGTGATCTAGTGTGTGCATGGTTTCCAG  
AACTGTCCATTGTAAATTTGACATCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTTATTAAGGTCCAGTGCA  
AAGCTTCTGTCTCCTTCTAGATGCATGAAATTTCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTTATT  
10 TCTTCTCGTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCAGTGGTAGAATTTTTATCTTCTGATGAGTGAA  
TCTTTTGGAGACTTCTATGTCTCTAGTAATCTAGTAATCTTTTTTAAATTGCTCTTAGTACTGCCACACTGGGCTTCT  
TTTGATTAGTATTTTCTGCTGTGCTGTTTTCTGCCTTTAAATTTATATATATATATATATTTTTTTTTTTTGGAGACA  
GAGTCTTGGTCTGTGCGCCAGGGTGTGAGTGCAGTGGTGTGATCAGGTCAGTGTAACTTTTACCTTCTGGCCTGAGCCGT  
CCTCTCACCTCAGCCTCCTGAGTAGCTGGAACGTGCAGACACGCCCGCTACACCTGGCTAATTTTTAAATTTTTCTGGA  
15 GACAGGGTCTTGCTGTGTTGCCAGGCTGGTCTCAAACCTCTGGACTCAAGGGATCCATCTACCTCGGCTTCCCAAAGTG  
CTGAATTACAGGCATGAGCCACCATGTCTGGCTAATTTTCAACACTTTTATATTCTTATAGTGTGGGTATGCTCTGTTA  
ACAGCATGTAGGTGAATTTCAATCCAGTCTGACAGTCGTGTTTAACTGGATAACCTGATTATTTTCAATTTTTTGTGTC  
ACTAGAGACCCGCTGGTGACTCTGATTCTCCACTTGCTGTTGATGTCCTCGTTCCCTGTTTCTCACCACCTCTTG  
GGTTGCCATGTGCGTTTTCTGCGAGTGTGTGATCCTCTCGTTGCCCTCGGTCACTGGGCATTTGCTTTTATTCT  
20 CTTTGCTTAGTGTTACCCCTGATCTTTTTATTGTGCTGTTGTTGCTTTTGTATTAGAGACAGTCTCACTCTGTCAACCA  
GGCTGGAGTGAATGGCACAATCTCGGCTCACTGCAACCTCTGCCTCCTCGGTTCAAGCAGTTCTCATTCTCAACCTCA  
TGAGTAGCTGGGATTACAGGCGCCACCACCGCCTGGCTAATTTTTGTATTTTAGTAGAGATAGGCTTTCACCATGT  
TGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTGATCTGCGCCGCTTGGCCTCCACAGTGCTGGGATTACAGGTGCAA  
GCCACCGTGCCCGCATACCTTGATCTTTTAAATGAAGTCTGAAACATTGTACCTTGTCTGAGCAATAAGACCCCTT  
25 AGTGTATTTTAGCTCTGGCCACCCCCAGCCTGTGTGCTGTTTTCCCTGCTGACTTAGTTCTATCTCAGGCATCTTGACA  
CCCCACAAGCTAAGCATTATTAATATTGTTTTCCGTGTTGAGTGTCTGTAGCTTTGCCCCGCGCTGCTTTCTCTCC  
TTTGTTCCTGCTGTCTCTGTCTCAGGCGCGCGTCTGGGGTCCCTTCTTGTCTTTGCGTGGTCTTCTGTCTGTG  
TTATTGCTGGTAAACCCAGCTTTACCTGTGCTGGCCTCCATGGCATCTAGCGACGTCGGGGACCTCTGCTTATGATGC  
ACAGATGAAGATGTGGAGACTCACGAGGAGGGCGGTATCTTGGCCCGTGAGTGTCTGGAGCACCACGTGGCCAGCGTTC  
30 CTTAGCCAGTGAGTGACAGCAACGTCCGCTCGGCTGGGTTGAGCTGGAAAACCCAGGCATGTGGGGTCTGGTGGCT  
CCGCGGTGTGAGTTTGAATCGCGCAACCTGCGGTGTGGCGCCAGCTCTGACGGTGTGCTGGCGGGGAGTGTCTG  
CTTCTCCCTTCTGCTTGGGAACCAGGACAAAGGATGAGGCTCCGAGCCGTTGTGCGCCAAACAGGAGCATGACGTGAGCC  
ATGTGGATAATTTTAAATTTCTAGGCTGGGCGCGGTGGCTCACGCCTGTAATCCAGCACTTTGGGAGGCCAAGCGGG  
TGGATCACGAGGTGAGGAGGTGAGACCATCTGGCCAAACATGATGAAACCCCATCTGTACTAAAAACAAAAAATTAGC  
35 TGGGCGTGGTGGCGGGTGCTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAAATTGCTTGAACTGGGAGTTGGAA  
GTTGCAGTGAGCCGACATTGCACCACTGCCTCCAGCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAA  
AAAAAAAAAAATTTAGTAGCCACATTAAGAAAGTAAAAAGAAAAGTGAAATTAATGTAATAATAGATTTTACTGAA  
GCCCAGCATGTCCACACCTCATTTTTAGGGTGTATTGGTGGGAGCATCACTCACAGGACATTTGACATTTTTTGGAGC  
TTTGTCTGCGGGATCCCGTGTGTAGGTCCCGTGTGCGTGGCCATCTCGGCTGGACCTGCTGGGCTTCCCATGGCCATGGCT  
40 GTTGTACCAGATGGTGAGGTCCGGATGAGGTGCGCAGGCCCTCAGTGAGCTGGATGTGCAGTGTCCGGATGGTGACG  
TCTGGGATGAGGTGCGCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGAGGTGAGGGTGTGAGGTCTCCAG

5 G C C C T C G G T G A G C T G G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G  
T G T G G T G T C T G G A T G G T G C A G G T C A G G G G T G A G G T C T C C A G G C C C T C G G T A A G C T G G A G G T A T G G A G T C C G G A T G A T G C A  
G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T G G G G T G A G G T C A C C  
A G G C C C T G C G G T G A G C T G G G T G T G C G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G A C G G T G C C A G A C C A T G C  
10 G G T G A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T C T G G G G T G A G G T T G C C A G G C C C T G C T G T G A G T T G G A T G T G G G G T  
G T C C G G A T G C T G C A G G T C C G G T G T G A G G T C A C C A G G C C C T G C T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T  
G G G G T G A A G G T C G C C A G G C C C T G C T T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G  
G C C C T C G G T G A G C T G G A T G T G C A G T G T C C A G A T G G T G C A G G T C C G G G T G A G G T C G C C A G A C C C T G C G G T G A G C T G G A T G  
T G C G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G G C C C T C G G T G A G C T G G A T G T A T G G A G T C C G G A T G G T G C C  
15 G G T C C G G G T G A G G T C G C C A G A C C C T G C T G T G A G C T G G A T G T G C G G T G T C T G G A T G G T A C A G G T C T G G A G T G A G G T C G C C  
A G A C C C T G C T G T G A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T C A G G G T G A G G T C T C C A G G C C C T C G G T G A G C T G G A  
G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A A C T G G A T G T G C G G C G T C T G G A T G G T  
G C A G G T C T G G G G T G T G G T C G C C A G G C C C T C G G T G A G C T G G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G G T G A G G T C G  
C C A G G C C C T G C T G T G A G C T G G A T G T G C G G C G T C T G G A T G G T G C A G G T C T G G G G T G T G G T C G C C A G G C C C T C G G T G A G C T G  
20 G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T T G C C A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G  
G T G C A G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C T G G G G T G A G G T  
C A C C A G G C C C T C G G T G A G C T G G T T G T G C G G T G T C C G G T G T G C A G G T C C G G G G T G A G T T C G C C A G G C C C T C G G T G A G C  
T G G A T G T G C G G T G T C C C C G T G T C C G G A T G G T G C A G G T C C A G G G T G A G G T C G C T A G G C C C T T G G T G G G C T G G A T G T G C C G T  
G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C A G G C C T T T G G T G A G C T G G A T G T G C G G T G T C T G C A T G G T G C A G G T C T G  
25 G G G T G A G G T C G C C A G G C C C T T G G T G G G C T G G A T G T G T G G T G T C C G G A T G G T G C A G G T C C G G C G T G A G G T C G C C A G G C C C T  
G C T G T G A G C T G G A T G T G C G G T G T C T G G A T G G T G C A G G T C C G G G T G A G G T A G C C A A G G C C T T C G G T G A G C T G G A T G T G G G  
G T G T C C G G A T G G T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C G G T T A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T  
C C G G G G T G A G G T C A C C A G G C C C T G C G G T T A G C T G G A T G T G C G G T G T C T G G A T G G T G C A G G T C C G G G G T G A G G T C G C C A G G  
C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C C G G G G T G A G G T C G C C A G G C C C T G C A G T G A G C T G G A T G  
30 T G C T G T A T C C G G A T G G T G C A G G T C T G G C G T G A G G T C G C C A G G C C C T G C G G T T A G C T G G A T A T G C G G T G T C G G A T G G T G C A  
G G T C C G G G T G A G G T C A C C A G G C C C T G C G G T T A G C T G G A T G T G C G G T G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C  
A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C C G G G G T G A G G T C G C C A G G C C C T G C G G T G A G C T G G  
A T G T G C T G T A T C C G G A T G G T G C A G G T C T G G C G T G A G G T C G C C A G G C C C T G C G G T G A G C T G G A T G T G C A G T G T A C G G A T G G  
T G C A G G T C C G G G G T G A G G T C G C C A G G C C C T G C G G T G G G C T G T A T G T G T G T G T C T G G A T G G T G C A G G T C C G G G G T G A G T T  
35 C G C C A G G C C C T G C G G T G A G C T G G A T G T G T G G T G T C T G G A T G C T G C A G G T C C G G G G T G A G T T C G C C A G G C C C T C G G T G A G C  
T G G A T A T G C G G T G T C C C C G T G T C C G A A T G G T G C A G G T C C A G G G T G A G G T C G C C A G G C C C T T G G T G G G C T G G A T G T G C C G T  
G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C A G G C C C T T G G T G A G C T G G A T G T G C G G T G T C C G G A T G G T G C A G G T C C G  
G G G T G A G G T C A C C A G G C C C T C G G T G A T C T G G A T G T G G C A T G T C C T T C T C G T T T A A G

## 35 Intron 3 (SEQ ID NO 6)

40 G T A C T G T A T C C C C A C G C C A G G C C T C T G C T T C T C G A A G T C C T G G A A C A C C A G C C C G G C C T C A G C A T G C G C C T G T C T C C A C T  
T G C C T G T G C T T C C C T G G C T G T G C A G C T C T G G G C T G G G A G C C A G G G G C C C C G T C A C A G G C C T G G T C C A A G T G G A T T C T G T G  
C A A G G C T C T G A C T G C C T G G A G C T C A C G T T C T C T T A C T T G T A A A A T C A G G A G T T T G T G C C A A G T G G T C T C T A G G G T T T G T A  
A A G C A G A A G G G A T T T A A A T T A G A T G G A A C A C T A C C A C T A G C C C T C C T T G C C T T C C C T G G G A T G T G G G T C T G A T T C T C T C  
T C T C T T T T T T T T C T T T T T G A G A T G G A G T C T C A C T C T G T G C C C A G G C T G G A G T G C A G T G G C A T A A T C T T G G C T C A C T

5 GCAACCTCCACCTCCTGGGTTTAAGCGATTACCAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC  
GCCTGGCTAATTTTGTACTTTTAGGAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG  
TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTAATT  
CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTAGGTATGAGAGGATAAAATCCACCCACTTGGCGACTCACTG  
10 CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTGAATG  
GCTGTGAGATTTTGTCTGCAATGTTCCGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTGCAATCAGTGAG  
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCCGCCAGGC  
TGACTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCAGCTCCCCTGCACACTCGAGTCCCTGGGGGGCCTTGTGAC  
ACCCCATGCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG  
15 GACCCCGACGTGGTGTCTGGGGCCATTTCTTGCATCTGGGGAGGGTCAGGGCTTCCCTGTGGGAACAAGTTAATACAC  
AATGCACCTTACTTAGACTTTACACGTATTTAATGGTGTGCGACCCAACATGGTCATTTGACCAGTATTTTGGAAAGAAT  
TTAATTGGGGTGACCGGAAGGAGCAGACAGACGTGGTGGTCCCAAGATGCTCCTTGTCACTACTGGGACTGTTGTTCTG  
CCTGGGGGGCCTTGGAGGCCCTCTCCCTGGACAGGGTACCGTGCCCTTTCTACTCTGCTGGGCCTGCGGCCTGCGGT  
AGGGCACCAGCTCCGGAGCACCCGGGCCCCAGTGTCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCACAGT  
20 GTGGCCGCTCCAGCCCCCGTGCCCCATGGGTGGTTTTGGGGAAAGGCCAAGGGCAGAGGTGTGAGGAGTGGTGGG  
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCCGCCCTCTCCATCTGAAGGATGTGGCT  
CTTTCTACCTGGGGTCTGCTGGGGCCAGCCTTGGGTACCCAGTGGCTGTACCAGAGGGACAGGCATCCTGTGTGG  
AGGGGCATGGGTTACGTGGGCCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGTGATGGTGGGACAGTACCCTGGGG  
GTTGACCGCGGACTGGGCGTCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG  
25 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCGGGCATGGCCTTCAGCGTGTGCTGCGTGGGTGCCCTGAGCCCTCAC  
TGAGTCGGTGGGGCTTGTGGCTTCCGTGAGCTTCCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT  
CTATTGCAG

#### Intron 4 (SEQ ID NO 7)

25 GTGGCTGTGCTTTGGTTTAACCTTCTTTTAAACAGAAGTGCCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG  
CCCGGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCACGGCGCCAACCCATTGTGCGCACAGTAGGTTGGCCGAGG  
TGCCGGTGCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG  
CAGCCGGGCCAGGGCCTGGATGCAGCACGGCCGAGGTCTGGATCCGTGTCCTGCTGTGGTGCGCAGCCTCCGTGCGCT  
30 TCCGCTTACGGGGCCCGGGGACCAGGCCAGGCTGCCAGGAGCCACCGGGCTCTGAGGATCTGGACCTTGCCCCACGG  
CTCTGCACCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCGCTCATCTGAGGAGAGTGTGGGGTGAGGTGGACAGAG  
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG  
TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCCTGGGTCCCTATGTTGGGGTGGGCAC  
TTGGCCGGATCCACTTTCCTGACTGTCTCCATGCTGTCCCGCCAG

#### 35 Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGACCCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG  
GAGGTACTCCTGGGTGGGCCGAGGGAGTGACGGTGACCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC  
CTTCAGCCTTTCTGCAGCACATGGGGCCGACTGTGCACCTGACTGCCCGGGCTCCTATTCCCAAGGAGGGTCCCACTG  
GATTCCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCCGGTGCTTGCACCCAGTCTGTAGCCAG  
40 GGGTCTCCTGCTCCTGAGGCTCAGAGAGGGGACACAGCCCGCCCTGCCCTTGGGGTCTGGAGTGGTGGGGTTCAGAGAGAG

5'-Bereich Intron 6 (SEQ ID NO 9)

15

TGTGGGATTGGTTTTTCATGTGTGGGATAGGTGGGATCTGTGGGATTGGTTTTTATGAGTGGGGTAACACAGAGTTCAAG  
GCGAGCTTTCTTCTCTGTAGTGGGCTCTGCAGGTGCTCCACAGCTTTATTGAGGAGACCATATCTTCTTTGAACATATGGT  
CGGGTTTATAGTAAGTCAGGGGTGTGGAGGCTCCCTGGGCTCCCTGTTCTGTTTCTTCCACTCTGGGGTCGTGTGGTG  
CCTGCTGTGGTGTGTGGCCGGTGGGCAGGGCTTCCAGGCCTCCTTGTGTTTCATTGGCCTGGATGTGGCCCTGGCTACGCT  
CCGTCTTGGAAATCCCCTGCGAGTTGGAGGCTTTCTTTCTTTCTTTTTTCTTTCTTTTTTTTTTTTTTGATAACAGA  
GTCTCGCTCTTTTTTGCCAGGCTGGAGTGGTTTGGCGTGATCTGGCTCACTGCAACCTGTGCTTCTGAGTTCAGCA  
ATTCTCTTGCCCTCAGCCTCCCAAGTAGCTGGAAATATAGGCGCCACCACCATGCTGACTAATTTTTGTAAATTTAGTAG  
AGACGAGGTTTCTCCATGTTGGCCAGGCTGGTCTCGAATCTCGACCTCAGGTGATCCTCCCACCTCGGCCTCCCAAAGT  
GCTGGGATGACAGGTGTGAACCGCGCGCCGCGGAGACTCGCTTCTGCACTTCCGTGAGATCTGCAGCGATAGCTG  
CCTGCAGCCTTGGTGTGCAACCTCCGTTTTCTTCTCCAGGTCTCGCTAGGGGTCTTTCATTTCATGACTCTCTTCA  
CAGAAGAGTTTACAGTGTGCTGATTTCCCGCTGTTTCTGCGTAATTGGTGTCTGCTGTTTATCGATGGCCTCCTTCCA  
TTTTCTTTAGGCTTTGTTTATTGTTGTTTTTCCGGCTCCTTGAAGGAAAAGTTTCGATTATGGATGTTTGAACTTTCTTT  
TCTAAACAAGCATCTGAAGTTGCCGTTTTTCCCTCTAAAGCAGGGATCCCAGGCCCTGGCTGTGGAGTGGCACCGGTCT  
GGGGCCTGTTAGGAACCGCGCACAGCGGAGGCTAGGTGGGGTGTGGGAGCCAGCGTTCCCGCCTGAGCCCGCCCC  
TCTCAGATCAGCAGTGGCATGCGGTGCTCAGAGGCGCACACCCCTACTGAGAACTGTGCGTGAGAGGGGTCTAGATTCT  
GTGCTCCTTATGGGAATCTAATGCCTGATGATCTGAGGTGGAACCGTTTGCTCCCAAACCATCCCCTTCCCCACTGCTG  
TCCTGTGAAAAAATCGTCTTCCACGAAACCACTCCCTGGTACCACAATGGTTGGGGACCTGTGCTAAAGACCTGCTTCA  
GCAGCCTCTCGTCAGTGTGATATATTGGCTTTTCTGTGTTGAGTCCAGAATAATTACGGATTCTCTGTGATGCTTCCGC  
CGACCTCAGACCATGGGCTATTGTGGGCGTGTGCTGCTCCTGGGTGGGAAGGGTGAGGCCCATGTACCTTCTCT  
GTTACTGCCTTCCAGGTTGGTTCTCAGGTTGAATCGTACTCGATGTGGTTTAGCCACGGCCCTGCCGCCAGCTCCTG  
GGGGCTGGGAACATGCTGAAGCACAGAGTCACCGTGGCGTCTTTTGTATGCCTCACAAGCTCGAGGCCCTCCTGTGTCG  
TGTTAGTGTGTGTACGTGCCGTGCTCACATCCTGTCTTGGGACGACAGGGGCTTAGCAGGTCCCGTAGTAAATGACAAGC  
GTCCTGGGGGAGTCTGCAGAATAGGAGGTGGGGGTGCCGGTCTCTCTCCCGCGTCTTCAGACTCTTCTCCTGCCGTGCT

40

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGGAGCTCGAGTGCCACTTGTGCCACGT  
GACTGTGGATGGCAGTCGGTCACGGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTGGTTCACAGGGGTCTGATGTGTG  
GTGACTGTGGATGGCGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTG  
ATGGCGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
5 GGGTCTGATGTGGTGAAGTGTGGATGGCAGTCGGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
TGGTGAAGTGTGGATGGCAGTCGGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
GTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
CGGTCTGATGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCA  
CAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCA  
10 GGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
TCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
GATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
GACTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGT  
GGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
15 TCGTGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
TCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
GGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
GACTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
GGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
20 GTGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
GGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
GTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCTGATGTGTGGTGAAGT  
ACTTTGCGTCTCGGCCCCCGGCCCCGTTTCCCAAACAAGCTTCCAGGCGCTCTCTGGGCTTCATCCCGCATCG  
GGCTTGGCCGAGGTCCACAGTCTGATCGGAAGAAACAAGTGCCAGCTCTGGCCGGGCGAGCCACATTGTGGCTC  
25 ATGCCCTCTCTGCGGCAG

#### Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGAAGTCCAGCAGTGGGTCTCCCTGGGCAATCACTGGGCTCATGACCG  
GACAGACTGTTGGCCCTGGGGGAGTGGGGGAATGAGCTGTGATGGGGCATGATGAGCTGTGTGCTTGGCGAAATC  
30 TGAGCTGGGCCATGCCAGGCTGCGACAGCTGCTGCATTAGGCACCTGCTCACGTTGACTGCGCGGCTCTCTCCAGTT  
CCGAGTGCCTTTGTTTATGATTTGCTAAATGCTTCTCTGCCAGTTTGTATCTTGAGGCCAAAGGAAAGGTGTCCCT  
CCTTTAGGAGGGCAGGCCATGTTTGGAGCGGTCTGCTGCCAGCTGGCCCTCAGTGTGGGTCTGAGGCCAAAGGAAAGC  
TGTCCCTCTCTTAGGAGGACGGGCGGTGTTTGGAGCCAGCCCCGCTGAGCGGGCTCTCAGTGTGGGTCTGTCCACGT  
GGCCCTGTGGCCCTTTGCAGATGTGGTCTGTCCAGTGGCCCTGTGGCTCTTTCAGATGCCTGTTAGCACTTGTCCGGC  
35 TCTAGGGGACAGTCTGTCCACCGCATGAGGCTCAGAGACCTCTGGGCGAATTTCTTGGCTCCAGGGTGGGGTGGAG  
GTGGCTGGGCTGTGGGACCCAGACCTGTGCCCCGCGAGCTGGGCGCAACTCTGGATCACATATGCCATCCGGGCCA  
CGGTGGGCTGTGTGGGTGTGAGCCAGCTGGACCCACAGGTGGCCAGAGGAGACGTTCTGTGTCCACACTCTGCCTAA  
GCCCATGTGTGTCTGAGAGACTCGGCCCGGCCAGCCACAGTGGCCCTGCATTCCAGCCAGCCCGCACTTCATACA  
AACACTGACCCCAAAGGGACGGAGGGTCTTGGCCAGTGGTCTGCTGTCTCAGCACCCACCGGCTCACTCCCATGTG  
40 TCTCCCGTCTGCTTTCGAG

## Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGCGGGTGGCTGGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCCTGCCC  
CTTCCCCACTGNCCTTCTGCCCCGGGGCCACCAGAGTCTCCTTTTCTGGCCCCCGCCCCCTCCGGCTCCTGGGCTGCAGGC  
5 TCCCGAGGCCCCGAAACATGGCTCGGCTTGCAGGAGCGGAGCAGGTGCCACACGAGGCCTGGAAATGGCAAGC  
GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGTGTGTCTGGGTGAGGTGTGCGCCGAGCGTTTGAGCCT  
GCAGCTTGTGAGCTCCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGATCTCTCTCTCCGATACAAAA  
GGATTTTATCCGATTCTCATTCTGTCCCTGTGCTGTGACCCCGCGAGGGCGGGGCTCTTCTCTGTGACTAGATTT  
CCCATCTGAAAAGTGGGGGTGACCGTGTAGTTTGTCTCCTCTCGGGGGGCTGTGGTGGCCATGGGGCAGGCGGCCTGG  
10 GAGAGCTGCCGTCACACAGCCACTGGGTGAGCCACACTCACGGTGGTAGAGCCACAGTGCCTGGTGCCACATCACGTCCT  
CTGGATTTTAAGTAAACACACACCTCCCGGCAGGCATCTGCCTGCGACCTGTGTGTGCTGGGAGAGTGGTAGCAC  
GGAGGAAATTCTGTGCACTCAAGGTCTATGCAAGGTCTCCGAGTCAGGTGGAACGTGGAGGCTCTCTCTGGGATC  
GTCTCCAGCGGATAAAGGACTGTGACAGCTTCGGAAGCTTTTATTTAAAAATATACTATTAATTATTGCATTATAAGT  
AATCACTAATGGTATCAGCAATTATAATTTTAAAGTATAATTAGAAATATTAAGTAGTACACAGCTTCTGGAAAAA  
15 CACAAATTGCACATGGCAGCAGAGTGAATTTGGCCGAGGGACAGTGTGCACATGTGTGTAAGCGGCCCCAGGCCCCAC  
AGAATTCGCTGACAAAGTCACTCCCCAGAGAAGCCACCAGGGCCTCCTTCGTGGTCTGTAATTTTATTAAGATGGATC  
AAGTCACGTACCGTCCACGTGTGGCAGGGCTTTGGGGAATGTGAGGTGATGACTGCGTCTCATGCCCTGACAGACAGGA  
GGTGACTGTGTCTGTCTGTCCCTAGGACACGGACAGGCCCCAAGCTCTAGTCCCATCGTGGTCCAGTTTGGCCTCTGA  
ATAAAACGCTCTTCAAACTGTGCCCCAAAACTAAGAACAGAGAGAGTTCCCATCCCATGTGCTCACAGGGGCGTA  
20 TCTGCTTGCGTTGACTCGCTGGGCTGGCCGACTCCTAGAGTTGGTGCCTGTGCTTCTGTGCAAAAAGTGCAGTCTCTT  
GCCCCATCACTGTGATATCTGCACCAGCAAGGAAAGCCTCTTTCTTTTCTTTCTTTTCTTTTGGAGACGGAACGTCA  
CTGTTGTCTGCCTGGGCTTGAAGTGCAGTGGCGCGATCTCAACTCACTGCAACCTCCGCTCCCGGGTTCAGCATTTTCTC  
CTGCCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCTGGCTAATTTTGTATTTTATAGTAGAGAG  
GGGTTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCTCCCAAAGTGCTG  
25 GGATTACAGGTGTGAGCCATCACGCCAGCCGAAAGCCTCTTTTAAAGTGACCACTATAGCGCTTCCCGAAAATAAC  
AGGTCTTGTGTTTTCAGTAGGCTGCAAGCGTCTCTTAGCAACAGGAGTGGCGTCTGTGGGCTCTGGGATGGCTGAGGG  
TCGCGTGGCAGCCATGCCTTCTGTGTGCACCTTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTTGTCTGAAAACGCA  
CCCTTGGCATCCTTGTGTTGAGAGTTTCTGCTTCTCGTTGGTCACTGAAACTAGGGGCAAGGTTGTATCCGTTGGCGC  
GCAGCGGCTACATGTAGGGTCATGAGTCTTTCACCGTGGACAAATCCTTGAAAAAAGAGAGTCCGGTTAAGCAT  
30 TCATTCGGGTCAAGTGTCTGGTTCTGTGAATAAATCTAAGATTTAAGAAACCTTAATGAAGAAACCTTGATGATTC  
AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTTCAGCCGCCCCAGTGCATGGTGAGAGTGGGAGCAGGGATTG  
TTTGTTCAGAGGTCTCATCTGGTATGTTCTGAGGTGTTTGGCCGGCTGAATGGTAGACGTGTGCTTGTGTGTATGAGGT  
TCTGTGTCTGTGTGGCTCGGTTTGAAGTGTACGCATGTCCAGCACATGCCCTGCCCGTCTCTCACCTGTGTCTTCCCGC  
CCCAG  
35

## Intron 9 (SEQ ID NO 13)

GTGAGGCCTCCTCTTCCCCAGGGGGGCTTGGGTGGGGGTGATTGCTTTTGATGCATTCAAGTGTAAATATTCTGGTGC  
TCTGGAGACCATGACTGCTCTGTCTTGAGGAACAGACAAGGTGACGCCCTTCTTGGTATGAAGCCGACGGGAGGGG  
TTGCACAGCCTGAGGACTGCGGGCTCCACGCAAGGCTCTGTCCAGCGGCCATGTCCAGAGGCCTCAGGGCTCAGCAGGCGG  
40 GAGGGCCGCTGCCCTGCATGATGAGCATGTGAATTCAACACCGAGGAAGCACACCAGCTTCTGTACGTCAACCAGGTTT

CGTTAGGGTCCTTGGGGAGATGGGGCTGGTGAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA  
CTGGGCGCCTCTTCAGCCCATTTGCCATCCCACTTGCAATGGGGTCTACACCAAGGACGCACACACCTAAATATCGTGCC  
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTAATACTTTAAGTTCTAGGGTAC  
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGTGTCACCCATTAACTCATCATTTACA  
5 TTAGGTATATCTCCTAATGTATCCCTCCCCACTCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACCCTG  
TGCCAAGTGTTCTCATTGTTCAAGTCCCACCTGTGAGTGAGAACATGTGGTGTGTTGGTTTTCTTTCTCTGCAATAGTTT  
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCTTTTTTATGACTGCATAGTATT  
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTGGTTGCAAGTCTTTGCTACT  
GTGAATAGTGCCGCAATAAACATACGTGTGCATGTGTTTATAGCAGCATGATTATAATCCTTTGGGTATATACCCAG  
10 TAATGGGATGGCTGGGTCAAATGGTATTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCCACAATGGTTGAA  
CTAGTTTACACTCCCAACAGTGTAAGTGTTCTGGTGCTGGAGAGGATGTGGACAGCAGTTATTTTTTTATGAAAA  
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATCGCTCAGGAAGCCTGCAGGCCACACAGCCATTTCTCTCGAAGAC  
TCCGGGTTTTCTCTGTGCATCTTTGAAACTCTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA  
GGTTCAAGTTCTAGATTGAATAAGTTTATGTAAACAGAAACAAAAATTTCTTGTAACACAACCTTGCTCTGGGATTTGGA  
15 GGAAAGTGCTCTGAGCTGGCGGCACACTGGTCAGCCCTCTGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG  
GCTTGGGCTGAGGGTCACACAGTGACCATGCCAGCTTCCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC  
ACAGCTGCCATGCTGGTAAAGGGCACCAGTGGCTCAGAGGGGGCAGGTTCCAGCCCCAGCTTTCTTACCGTCTTCAG  
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCCTTCGTTCTGCTTCAGCTGGCACAGAATTGCACAA  
GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTGGAAA  
20 GAAATTTAAGTTTTTCAATTAACCGCTTGGAGAATGTTACTTTATTATGGCTGTGTAATTTGTTGACATTAGTCCC  
TCGTAGACAGATACTACGTAAAAAGTGTAAGTTAACCTTGCTGTGATTTTCCCTTATTTTAG

#### Intron 10 (SEQ ID NO 14)

GTGAGGCCCGTGCCGTGTGTCTGTGGGACCTCCACAGCCTGTGGGCTTTGCAGTTGAGCCCCCGTGTCTGCCCCCTGG  
25 CACCGCAGCGTTGTCTCTGCGCAAGTCTCTCTCTGCGGCTGTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC  
CAGGCTCGGGGCGCAGGGGACCTTCGGGAGGGAGTGGGTACCGTGACGGCCCTGGTCTGACAGAGCGCACCCAGGTT  
ACACAGTGGTGAGTGACGGCGGTGACCTGGCTCCTGCTGCTCTTGGAAAGTCAAGAGTGGCGGCTCCTGGGGCCCCAG  
TGAGACCCCGAGGAGCTGTGCACAGGGCTGCAGGGCCGAGGCGGCAGCCTCCTCCCGAGGTCACCTGAGCCTGCGGA  
GAGCAGGAGCTGCTGAGTGAGCTGGCCACAGCGTTCTGCTGCGGTACGTTCTGCGTGGGGTTGTTTGGGATCGGTGGG  
30 AGAATTTGGATTTGCTGAGTGCTGCTGTTGAACCACGGAGATGGCTAGGAGTGGGTTTCAGAGTTGATTTTGTGAAT  
CAAATAAAATCAGGCACAGGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG  
AGCCGGTGGGCTTGTTTTAAAGTGCGATTGACGAGGGACGAGAAACCTTGAAAGCTGTAAAGGAACCTCAGAAAATG  
TGGCCGCCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTTGTGAAAACCATTTTGACCCGCCCTCCAAGTCCACCC  
TCCAGGTCCACCTCCAGGGCCGCCCTGGGCTGGGGATGACCTGGCGTTCTTGTCGCCGAGCCCGAGCACAGCAGGC  
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT  
GCTGAGGGGACAGAGCAGACGGGGAACGCTGCTTCTGTGTGGCAAGTTCCTGAGGGTGTGGCCAGGGAGGTGGCTCAGA  
GTGTATGTTGGGTCCTACCGGGGGCAGAACTCTGTCTCTGATGAGTCGGCAGCCATGTAACAGGAAGGGGTGGCCACAG  
GGAGCTGGGAATGACACAGGGGAGCTGCGCAGCTGGCCGAGGTCCAGGGCCAGGCCACAGGAAGGGCAGGGGGACGCCC  
GGGGCCACAGCAGAGGCCCGCAGGAAGGGAAGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGCTCCCTG  
40 AGCTGGGTGAGCGAGGCTCATGACTCGGCGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGTCACAGCTCACAG

5 CCCAGCCAGGTCCCGCGCCTGAGCAGGAACTCAGAACCCTCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT  
AGGAGAAAACAGGCAAAGTCGTTGAGAAACGTCCTTAAAAGAAGGTGGGATGGTGGCAATTTCTGTCCAGATTTTAGTCT  
GCCCCGACCACAGATGAGTCTATAACGGGATTGTGGTGTGCCATGGGGACACATGAGATGGACCATCACAGAGGCCAC  
TGGGGCTGCACCTCCCATCTGAGTCTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGCTATGCTACCTACCTGTCTCTGCC  
GGGAGACAGGGAAGCACCCGAAGCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCAGCACCTT  
GCTCCAAATCACCCTTCTCTGGGGTTTTCCAAAGCATTAACAAGGGTGTGAGGTTACCTCCTGGGTGACGGCCCCGCA  
TCTTGGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGGTGGGGCAGGTGCTGCTGCAGGGCCGTTGCGTCCACCTCTGCT  
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGTCAGAGGCCACAGGGTGCCCTCGTCCCATCTGGGGCTGAGCAGA  
AATGCATCTTTCTGTGGGAGTGAGGGTGCTCACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAAAGGAAATGGTGAC  
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAAGACGCCCGGGCCCTGCTGGGCGTGAGT  
CTCTCAAACCGAACACAGGGGCCCTGCTGGGCATGAGTCCCTCTGAACCCGAGACCTGGGGCCCTGCTGGGCGTGAGT  
15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC  
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA  
AATTCTGGGGTCTTGTTCCTCCAGAGCCGAGAGCTCAAGGCCCGCTCTCAGGCTCAGACACAAATGAATTGAAGATGGA  
CACAGATGCAGAAATCTGTGCTGTTTCTTTATGAATAAAAAGTATCAACATTCAGGCAGGGCAAGGTGGCTCACACCT  
ATAATCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTGAGGCCAACCTAACCAACATAGTG  
20 AAATTCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGGCACACGCCTGTAGTCCCCGCTATGCGGGAGGC  
TGAGGCAGGAGAATCATTGAACCCAGGAGGCAGAGGTGTCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCA  
ACAGAGTGAGACTTCATCTTAAAAAAGAGTATCAGCATTCAAAACCATAGTGAGCAGGTGTTTTTTTATTC  
TGTCCTTCGATAATATTTACTGGTGCTGTGCTAGAGGCCGGAACCTGGGGTGCCCTCCTCTGAAAGGCACACCTTCATGG  
GAAGAGAAATAAGTGGTGAATGGTTGTAAACCAGAGGTTTAACTGGGGTCTGTGCTCTGAGTTAACAGTCCAGATC  
25 TGGACTTTGCCTCTTTCCAGATGCTCCCTGGGGTTTGCTTCATGGGGAGCAGCAGGTGTGGACACCCTCGTGATGGGG  
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGAGTGGCAGGTGCAGACACCCTTGTCATGGTGGCCAGCATGCTCCCTG  
TTGCAGCTCCCTCCCCACAAGGATGCCGGTCTCCTGTGCTCCCCACAGTCCCTGCTTCCCTCTCACAGCCTTACCTGGTC  
CTGGCCTCCACTGGCTTTGTCTGCATGATTTCCACATTTCTGGGCTCCACAGCCTCTTCGCTCTCCAGGCACCTCT  
GCAGTGCTGGCCATACCACTGAGTGTGAAGTGTCCACTGCTTATTTTGTCTCCCATGAAATGTATTTTATAGGACAGGC  
30 ACCCTGGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAAATGG  
GTTCTCTCTAAACACATTGCAAAGCCACAGAGGCTAGTGCAGGATGGGTGGGCATCAGGTATCAGATGTGGGTCCAATG  
CCAGAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTGTCAGAGGTGGCTCTAAAAGCTCAGCAGTGGAG  
GCAGTGGTTCCGCATACTCAGGGTGAACCTCACATCCTCTGTGTCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA  
GAAGAAAACAGGCAAAATGATTAAGAAAAGTGAAGGAAAGTGGTAAGATGGGAATTTCTTGTCCAGATTTTAGTCT  
35 TCCCAAACCAAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAAACATAGAACAAACGGAAGCCCTATCTCT  
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAGAGAGTGTGTGTGTAATTTTTTTTCTGAGAAAAC  
GACTGGAAGCAAAATAAGTTGTGCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACATGCAAAACAC  
ACCGAACACAGAAATAAAACAAAAGACTCAAAGGGAAGGGAGGTGAACGTTCCCTGGTTTGGTGTGGGGAAGGACACAC  
AGGGAGGCGGATGAAACCACTGAGGCAACGGGCATTGCTTTCACTGCAGAGAACTCAGCTTGCTGAGCCACAGTGAAA  
40 ATGGCCATTCCCTGGAGCGTTTGTGACGTGATTTATTTAAGGCGCCCTGTGAGGTCCTGCACATTCATCTCTCACTTT



5 GTTCTCCTAACACCTGAGAGGTAGAGGAGGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGGC  
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCCGAGGACCCACACAAGTCAGACCCATAGGCTC  
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCGTATGCACACTTGGGAAGGTC  
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC  
10 GAGACCTGTCCCATCCCTCATGCTGGCTCCTTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG  
ACTTTTCTGGAAAGCAGCTTGTTCATGGAAGTCCTCACAATGTCTGTGCTTCCAGTAATTCACCTTCTGAAGTGA  
CCAGACATTATCACGGGTCTTATTTACCATTTCAGTGTTCCAGGCAGGGGACTTGCCACAGCAAGTCACGAACCTGCC  
CAAATACAGGGCTAAGGAGATATTATGCATCAGAAACTTGCTCTGCCATTAAACATTTTCAAAGAATTTTGAAGAAT  
GTTTAATGGCACAAAACGTTTATTTCATGTAGCAGTGTTCAAAGCTGGATGTAAAGAACAACACCCAGGAGCCTGCCG  
15 TGAATGTATGTGTGTTCATCTTTGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG  
GATGCCTCCATCCTGCCCTCTGGAGACCATGTGTGCCAGTGCACTCACTGGAGCCCTGTTAGCTGGTGCCACCTG  
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCCAACTCAGTGTTCTCCCAAAAAACCTGAGTCAC  
ACCTGTGTTCACTCGAGGGACGCCCGGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTTGGCTGAGTTATGTGCAGATC  
TCATCAGGGCAGATGATGAGTGCAAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCCTGGTG  
20 GAGTTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTTGGAGTCCATGGAGTGAGCACCAGCCCCCTCGGGCTGCAGC  
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTGGAGCAAGCTTTCAGGAGGGGGCTGGGTGT  
GGGCGAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

#### Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCGT  
GCCTGCAAGGCTGATGGTGA CTGGCTGCACGTAAGAGTGACATGTACGCATATACACGTGAGCACATACATGTGTGCAT  
GTGTGTACATGAAGGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCACATGCGAATGCACACCTGACA  
TGATGTGTGTTCTGTGCACAGTCGTGTGGGCATTACCGTGAGGTGCATGCGTGTGGGTGTGCAGTGTGAGTAGCATGTGT  
GCACATAACATGTATTGAGGGGTCTCGTGTTCACCCCGTAGTCTCTCAGCACCAGTGCCACTCCTTACAGGATGAGAC  
25 GGGGTCCAGGCCTTGGTGGGCTGAGGCTCTGAAGTGCAGCCCTGAGGGCATTGTCCCATCTGGGCATCCGCGTCCACT  
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTACATCCACTCCACTCCCTCTCTCTGTGGGC  
ATCCGCGTCCACTCCCCCTCTCTGTGGGCATCTGCGTCCACTCCCCCTCTCTGTGGGCATTGCGTCCACTCCCTCTCCT  
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGCAGGCAGATGACACAGAGTCTTGACTCGCCAGGGTGGTTTCGAGCTG  
CCGGGTGAGGGCCAGGCCGATTCTCACTGGGAAGAGGGATAGTTTCTTGTCAAATGTTCTTTCTTGTTCATCTGA  
30 ATGGATGTAAAGCAAAAAGTAAAACTTAAATCCAGAGAGGTTTCTACGTTTCTCACTCTTTCTTGGCGACTCTAG

#### Intron 13 (SEQ ID NO 17)

35 GTGAGCCGCCACCAAGGGGTGAGGCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACTT  
TGGAACCTCTGGGTTTTAGGGGCAAGGAATGTCTTACGTTTTTCACTGGTGCTGCTGCCTGTGCACAGTTCTGTTTCGCGTG  
GCTCTGTGCAAGCACCTGTTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCACTGTGCCTGT  
GCACTGGCCGTGGGACGTATGGAGGCATCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG  
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGAGGATTTGGGGTCTCAGCAAAGAGGGCC  
GAGGTGGGTGACAGGTGAGGGTCGCTGGCCCCACCCCGGAAGGTGCAGCAGAGCTGTGGCTCCCCACACAGCCCGGCCA  
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCCTGTCTGGCTGGTCAGGGGGTCCCTGCCAAGAATCG  
40 ACAACTTTATCACAGAGGGAAGGGCCAATCTGTGAGGCCACAGGGCCAGCTTCTGCCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAAGGGCAGTCGGGCACCACAGGCCCGGGCCTCCACCTCAACAGGCCTCCCGAGCCA CTG  
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTACCCAGGGCCG  
AGGCTGCGCGAATTACCGTGACACTTGATGTGAAATGAGGTCGTCTATCGTGAAACCCAGCAAGGGCTCACGGGA  
GAGTTTTCCATTACAAGGTCGTACCATGAAAATGGTTTTTAACCCGAGTGTGCGCCTTCATGCTCTGGCAGGGAGGGC  
5 AGAGCCA CAGCTGCATGTTACCGCCTTTGCACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCAGGGTGCCTCC  
GGCTCAGACCGCCCTCCTCTCTGCTTCTCTCTCTGCTCAAATCTTCCCTCGTTTGATCTCCCTGACGCGTGCTGGG  
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGAAACCCCTTGGGGTGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT  
CTGACACTGTGGTTGACCCAGGGTCCAGCTGGCGTGCTGGGGCCTCCTTGGGCCATGATGAGGTGAGGAGAGTTTTCC  
CAGGTGAAAACCTCCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTAGCTCAGTCTTGTCTCTC  
10 ATTTCCCA CACCAGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCCTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCCAC  
CCATGTGGGGACCCCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGCTTCCA  
AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTTCCCTCGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG  
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGGCGGGGTCTGATTCAAATCCGC  
TGGGGCTCGGCCTTCTTGGCCCCTGTGCGCCGCTCCACACGGGCTTGGGGTGGACGCCCGACCTCTAGCAGGTGGC  
15 TATTTCTCCCTTTGGAAGAGAGCCCTCACCCTAGCTAGGTGTTTCCCTCCTGGGTGAGGAGCGTGGCCGTGTGGCAACC  
CCGGGACCTTAGGCTTATTTATTTGTTTAAAAACATTCTGGGCTGGCTTCGTTGTTGCTAAATGGGGAAAAGACATCC  
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGTGTGGCTTGACTGGTGTGATCTCAGGTCAATCCAGAAGTGGCT  
CAGGAAGTCAGTGAGACCAGGTACATGGGGGCTCAGGCAGTGGGTGAGTGAGGTACACGGGGGCTCAGGCAGTGGGT  
GAGGCCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGAGGCTCAGACCAGGTAC  
20 ACGGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC  
CCCAAAGTCCAGGAAGCTGAGAGGCCAAAGATGGAGGCTGACAGGGCTGGCGGGTGGCTCACACCTGTAGTCCCAGCA  
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCAGGAGTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC  
TATGAAAAATAAAAAAAAATTAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG  
GATCACTTGAGCCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA  
25 GCCCATCTCAACAACAACAAGAAGACTGACAAATGCAGTTTCTTGAAAGAAACATTTAGTAGGAACCTTAACCTACACA  
CAGAAGCCAAGTCGGTGTCTCGGTGTGAGTGAGATGAGATGATGGGTCTCACACCATCACCCAGACCCAGGGTTTATG  
CACCACAGGGGGCGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGTTGTCTGACGAAGGGCAG  
GATTCATGATAAGTACCTGTGTTACACAAGGAACAATGGATAAACTGGAACCTTAGAGGCCTTCCCGAACAGGGGCT  
AATCAGAAGCCAGCATGGGGGGCTGGCATCCAGGATGGAGCTGCTTCAGCCTCCACATGCGTGTTCATACAGATGGTGCA  
30 CAGAAACGCAGGTACCTGTGCACACACAGACACGAGCTACTCGCACACACAAGCACACACAGACATGCATGCATGC  
ATCCGTGTGTGCACCTGTGCCCATGAGGAAACCATGTCATGTGCATTATGCACGCACACAGGCACCGGTGGGCCCAT  
GCCACACCCACGAGCACCGTCTGATTAGGAGGCCTTCTCTGACGCTGTCCGCCATCCTCTCAG

**Intron 14 (SEQ ID NO 18)**

GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCCTGCCTGCTGGTGTAGTGTGTCAGGAGACTGAGTGAATCTGGG  
35 CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCCAACCACTGTGAGGCTCGTCTGCCCCCCTCTCGT  
GGGGTGAAGCAGACACCTGATGGAAGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCACCTTGCTCTGCCTGGGGAA  
GCGCTGGGGGCGGTGGTCTCTCTGTTTGGCCCATGGTGGGATTGGGGGGCCTGGCCTCTCTGTTTGGCCTGTGGTGG  
GATTGGGCTGTCTCCGTCCATGGCACTTAGGGCCCTTGTGCAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG  
40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGCTATCACACGACAGAGCCCGCGCGCTCTGCTTCCAGTCACCG

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAAGCCATGTGAACTGCGGT  
CCTGAGCTTAACAGCTTCTACTTTCTGTTCTTTCTGTGTGTGGAAATTTACCTGGAGAAGCCGAAGAAAACATTCTCG  
TCGTGACTCCTGCGGTGCTTGGGTGCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTTCCG  
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG

5

**Intron 15 (WEQ ID NO 19)**

GCAAGTGTGGGTGGAGGCCAGTGCAGGGCCCCACCTGCCCAGGGGTCTCCTTGAACGCCCTGTGTGGGCGAGCAGCCTC  
AGATGCTGCTGAAGTGCAGACGCCCCCGGGCCTGACCTGGGGGCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG  
CTGGTGTCCCCAGGCCACGGAGCCTGGCAGGGTCCCCAACTTCTTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC  
10 CCAGCCTTGGGAGCCTTCTGACCCCTGACCTGTGTCTCTCACAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGGT  
CCTGAGCAAGTTCTCTCCCGCCCCGCGCTCCAGCGTCACTGGGCTGCCTGTCTGCTCGCCCGGTGGAGGGGTGTCTG  
TCCCTTCACTGAGGTTCCACAGCCAGGGCCACGAGGTGCAGGCCCTGCCTGCCCGGCCACCCACAGCTCTAGGAGGG  
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTTGGCCCCGAG

**3'-untranskribierter Bereich (SEQ ID NO 20)**

ATCTCATGTTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT  
GGTCAGTGCAGGCCCATGGCCTGGCTGTGCATTTACGGAAGTCTATGAGTGAATGGGGTTGTGGTCAGTGCAGGCCCATG  
GCCTGGCTGGGCTGGGAGGTTTCTGATGCTGTGAGGCAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA  
CCCTGGAAGACATAACAGTAAGTCCAGGCCCGAAGGGCAGCAGGGATGCTGGGGGCCAGCTTGGGCGGCGGGGATGATG  
20 GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGGCCACAGCTGGGGTGGCAGGGGTGATGGGGGGGGTGGTCTGGGTGG  
CGGGGAAGATGGGAAGCCTGGCTGGGCCCCCTCTCCCTGCCTCCACCTGCAGCCGTGGATCCGATGTGCTTCCCT  
GGTGACATCCTCTGGGCCATCAGCTTTTATGAGGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT  
CCTCCTCCTGAACGCCCAACTCAGGTTGAAAGTCAATTCGCGCTCTGGCCATTCTCTTAAGAGTAGACCAGGATTCTG  
ATCTCTGAAGGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTGTGCTCTCTC  
25 ATCTCTTATCATCTCCAGTCTCATCTCTCATCCTCTTATCATCTCCAGTCTCATCTGCTTCTCTTATCTCCAGT  
CTCATCTGTCTCCTCTTACCATCTCCAGTCTCATCTCTTATCCTCTTATCTCCTAGTCTCATCCAGACTTACCTCCA  
GGGCGGGTGCAGGCTCGCAGTGGAGCTGGACATACGTCTTCTCAGGCAGAAAGAACTGGAAGGATTGCAGAGAACAG  
GAGGGGCGGCTCAGAGGGACGCAGTCTTGGGGTGAAGAAACAGCCCCCTCTCAGAAGTTGGCTTGGGCCACACGAAACCG  
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCTACTGAGTG  
30 CACCTTGGACAGGGCTTCTGGTTTGAGTGCAGCCCGACGTGCCTGGTGTGGGGTGGGGCTTATGGCCACTGGATATG  
GCGTCATTTATTGCTGTGCTTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGCCAAGTCCACAGACTG  
TGTCGTAAATGCACTCTGGTGCCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTCTTGGCAGCCGGCTGGGG  
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGGCCCGGGCGCGCTGGGCGGACGACCTCAAGTGAGAGGTTGGACAGAAC  
AGGGCGGGGACTTCCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGTTTGAATCACAGACCAACAGGTCAGGCCATT  
35 GTTCAGCTATCCATCTTCTACAAAGCTCCAGATTCTGTTTCTCCGGGTGTTTTTTGTTGAAATTTTACTCAGGATTACT  
TATATTTTTTGCTAAAGTATTAGACCTTAAAAAAGGTATTTGCTTTGATATGGCTTAACCTACTAAGCACCTACTTTAT  
TTGTCTGTTTTTATTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCAACCCAGGTTGTTAGTGCAGTGGCAC  
AGTCATGGCTCGCTGTAGCCGCAACCCCCAGGCTCAAGTGATCCTCCGGCCTCAGCTTCCAGAGTGTGGGATTACAG  
GTGTGAGCCACTGCCCTTGCCTGGCACTTTTAAAAACCACTATGTAAGGTGAGGTCCAGTGGCTTCCACACCTGTCTATCC  
40 CAGTAGTTTGGGAAGCCGAGGCAGAGGATTGTCTGAGGCCAGGAGTTTGAAGCCAGCATGGGTAAACATAGGGAGACCCC

ATCTCTACAAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGTGTCTCGGGAGGCTGAGTGGG  
AGGATCGCTTGAGCCCGGAGGTGATGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA  
GACCTGTCTCAAAAAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGGAAGAAGGAAAGAGAAGAAGAAG  
GAAGAAGGAAGAAAGAAGGAGAAGGAGGCCCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAAAACA  
AAGTTTTAAAGGGAAGAAAAACCCAGCTCTTTGGACTTCCTTAGGCCTGAACCTCATCTCAAGCAGCTTCCTTCCACA  
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA  
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCTGCTCAGCCCCACGCTCCTGCCGGTCTGCACTGCTGTAACCGTC  
GATGTTGGTGCCAGGTGCCCACTGGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTGGTGGGTTTCAGAAGCCCCAG  
GCACTTGTGGCAGGCACAATTACAGCCCTCCCCAAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTCACCCG  
CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCATCTGGATTATCTGG  
TGGGCCTGATATGGCCACAAGGTCCTTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC  
CACTGGCCACTGCTGGCTTTGAGATGGAGGAGGGGGTCCCCAGCCAAGGAATGGGGGAGCCGCTCCATGCTGGA AAAAGC  
AAGCAATCCTCCCGGTCCTGAGGGCACACGGCCCTGCCACGCCCTCGATTTCAGGCCAGTGGGACCTGTTTCAGCTTTC  
CGGCCCTCAGAGCTGTAAGATGATGCGTTTGTGTTTCAGCCACTAAGCTGCAGTGATTCTGCACAGCAGCAAAATGGAATAG  
CAGTACAGGGAATGAATACAGGGACAGTTCTCAGAGTGACTCTCAGCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

15

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorenght. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

20

25

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

30

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmen aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

#### Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA<sup>+</sup>-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur  
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv  
markierter Primer (5'GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10<sup>5</sup>  
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl  
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),  
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl  
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser  
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die  
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach  
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden  
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl  
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-  
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min  
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol  
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stopplösung gelöst. Nach 5  
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel  
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom  
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in  
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA<sub>n</sub>TTGT)  
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA<sub>n</sub>Na/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losen Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei  
10 anderen TATA-losen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

#### Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem  
25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer  
30 zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer



angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCAGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

### Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

### Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGATTTCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII-Erkennungssequenz. Die  
5 verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und  
10 anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 $\alpha$ -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für  
15 die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

#### Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection  
25 bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

30

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO<sub>2</sub> kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit  
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-plattieren.

- Für die Transfektion wurden pro Well jeweils  $2 \times 10^5$  HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium  
15 entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium  
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H<sub>3</sub>PO<sub>4</sub>; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl  
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO<sub>3</sub>)<sub>4</sub> Mg(OH)<sub>2</sub>·5H<sub>2</sub>O; 2,67 mM MgSO<sub>4</sub>; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl<sub>2</sub>;

50 mM  $\beta$ -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100  $\mu$ l 1 M  $\text{Na}_2\text{CO}_3$  gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK  
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im  
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

## Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl. Acad. Sci.* 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). *Current protocols in molecular biology*. Greene Publishing Associates and Wiley-Intersciences, New York.
- Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. *Nature Genetics* 12, 200-204.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. *Proc. Natl. Acad. Sci.* 92, 9082-9086.
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* 11, 1921-1929.
- 20 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. *Science* 269, 1236-1241.
- 25 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. *Mol. Cell. Biol.* 14:4894.
- 30 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. *Science* 249, 1129-1133.
- 35 Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. *Nature* 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

- Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.
- 5 Shay, J. W. (1997). Telomerase and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.
- Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.
- 10 Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.
- 15 Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.



**Patentansprüche**

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit.
- 5 2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser Sequenzen handelt.
- 10 3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.
- 15 4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß  
20 es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide oder Proteine kodieren.
6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
- 25 7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß einem der Ansprüche 4 bis 6.

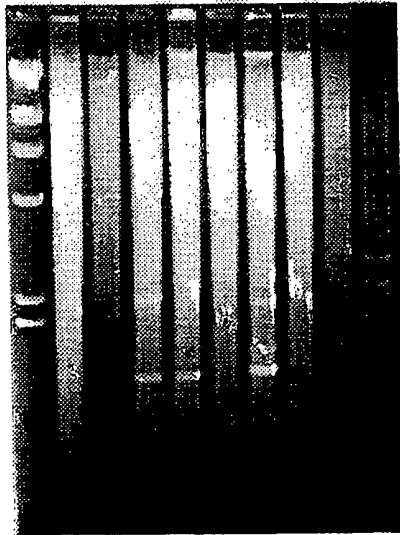
9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5           A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10
10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.
- 15
11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20
12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25           mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

A

1 2 3 4 5 6 7 8 9 10



B

1 2 3 4 5 6 7 8 9 10

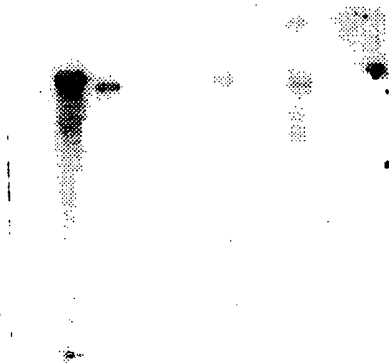
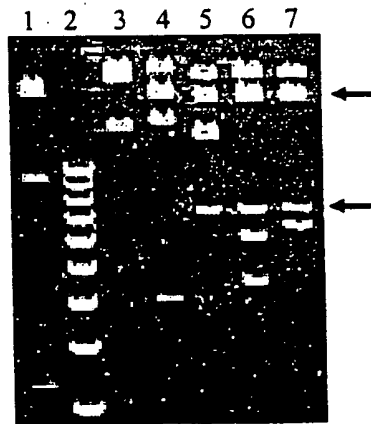


Fig. 2



ERSATZBLATT (REGEL 26)

Fig. 3

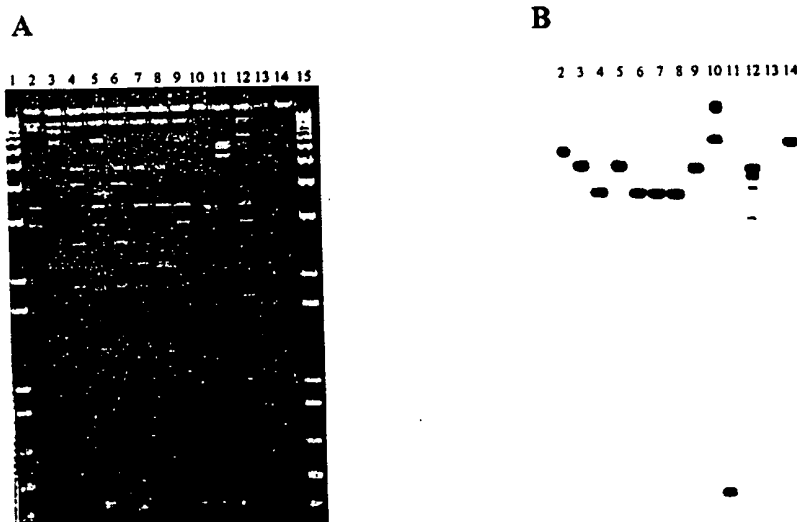


Fig. 4

GAGCTCTGAA CCGTGGAAC GAACATGACC CTTGCTCGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA 70  
 GTGGTGTGCA GGAAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATATT 140  
 CATCTTCACC CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAAACCTAG 210  
 TACAAACACC ACTCTTTTAC TAGGCCACCA GAGCACGGGC CACACCCCTG ATATATTAA AGTCCAGGAG 280  
 AGATGAGGCT GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTC TCTAGACTAG 350  
 TAGACCTGG CAGGCACTCC CCCAAATCT AGGGCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG 420  
 GAGACTCAGC CTGGGGTGCC ACACTGAGGC CAGCCCTGTC TCCACACCT CCGCTCCAG GCCTCAGCTT 490  
 CTCCAGCAGC TTCTAAACC CTGGGTGGGC CGTGTTCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT 560  
 TGCTCAGCG ACGTAGCTCG CACGGTTCCT CCTCACATGG GGTGTCTGTC TCCTTCCCCA ACACTCACAT 630  
 GCGTTGAAGG GAGGAGATTG TGGGCTCCG AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC 700  
 GATGCAGGTT CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCGCTCTC CTGTCACTG 770  
 CCGGGGCTCG CCGGTGTGTT CTCTGTTC TGTGCTCCTT TCCACGTCCA GCTGCGTGTG TCTCTGCCCG 840  
 CTAGGCTCTC GGGGTTTTTA TAGGCATAGG ACGGGGCGGT GGTGGGCCAG GCGGCTCTTG GGAATGCAA 910  
 CATTGGGTG TGAAGTAGG AGTGCTGTC CTCACCTAGG TCCACGGGCA CAGGCTGGG GATGGAGCCC 980  
 CCGCCAGGGA CCCGCCCTTC TCTGCCAGC ACTTCTCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT 1050  
 TTCCACAAGC ACTAAGCATC CTCTTCCCA AAGACCCAGC ATTGGCAGCC CTGGACATT GCCCCACAGC 1120  
 CCTGGGAATT CAGGTGACTA CGCACATCAT GTACACACT CCGTCCACGA CCGACCCCG CTGTTTTATT 1190  
 TTAATAGCTA CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAAC TGGTTAAACA AACGGGTCCA 1260  
 TCCGCACGGT GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAG 1330  
 GAATTACGCT GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAAGAATTTC 1400  
 ACCCATGGC AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA 1470  
 CCTTTTACTA AAGCCAGTTT CTTGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG 1540  
 TGGGGATGGG GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCGGGCAGG ATAATGCTCT 1610  
 AGAGATGCCC ACGTCTGAT TCCCCCAAAC CTGTGGACAG AACCCGCCCC GCGCCAGGGC CTTTGCAGGT 1680  
 GTGATCTCCG TGAGGACCTT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGGCCGAAAA GTAATCCAGG 1750  
 GGTCTCTGGA AGAGGGCGGC AGGAGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA 1820  
 GGCTGAAAG GGAGGGAGGG CCTCGAGCCC AGGCCTCCAG AAGCTGGAAA AAGCGGGGAA 1890  
 GGGACCTCC ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC 1960  
 TCCGGCTCC GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA 2030  
 GCAACAGGAA ACCCATGCAC TGTGAATCTA GATTATTTC GGCAGGCAG AGTGATTITA TTTAGCTATT TATTTTATT 2100  
 CAGGGCTGAA GTGCTCCGG GCAAGGGCAG GGCAGGCAG AGTGATTITA TTTAGCTATT TATTTTATT 2170  
 TACTTACTTT CTGAGACAGA GTTATGCTCT TGTGGCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA 2240  
 CTGCAACCTC CGTCTCTGG GTTCAAGCAA TTCTCGTCC TCAGCCTCCC AAGTAGCTGG GATTTCAGGC 2310  
 GTGCACACC ACACCCGCT AATTTTGAT TTTTAGTAGA GATGGGCTT CACCATGTTG GTCAAGCTGA 2380  
 TCTCAAAATC CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC 2450  
 CACTGCACCT GGCCTATTTA ACCATTITAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT 2520  
 CATGGAGTTC AATTCTCCCT TTACTCAGGA GTTACCCTCC TTTGATATT TCTGTAATT TCTGTAGCT 2590  
 GGGGATACAC CGTCTCTTGA CATATTACA GTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG 2660  
 GGGCAGCTGG GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA 2730  
 ATCAGGGCGC AAGTGTGCAG ACTGCTGGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA 2800  
 AAGTCCATCC CTCCTACTCT ACTGGGATTG AGCCCTTCC CTATCCCCC CCAGGGGCG AGGAGTTCCT 2870  
 CTCACTCTG TGGAGGAAGG AATGATATT TGTATTITT CACTGCTGGT ACTGAATCCA CTGTTTCATT 2940  
 TGTGTTTGG TTTGTTTTGT TTTGAGAGGC GGTTCACCTC TGTGTTCTCA GGTGAGAGG AGTGCATGG 3010  
 CCGGATCTTG GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCCG CTCCCATTTG 3080  
 GCTGGGATTA CAGGCACCCG CCACCATGCC CAGCTAAATT TTTGATTTT TAGTAGAGAC GGGGGTGGGT 3150  
 GGGGTTCCAC ATGTTGGCCA GGCTGGTCTC GAACCTCTGA CCTCAGATGA TCCACCTGCC TCTGCCCTCT 3220  
 AAAGTGCTGG GATTACAGGT GTGAGCCACC ATGCCAGCT CAGAATTTC TCTGTTAGA AACATCTGG 3290  
 TCTGAGGTAG GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCCAA TGATAGAAAT TTTTATTGT 3360  
 TGTTAGAACA CTCTTGATGT TTTACTGTG GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC 3430  
 TGCACCCATA AACTTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCTCTG 3500  
 CCATGCACAT GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTTCCCT CTTTAAAAAT 3570  
 TGTGTTTTCT ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAACT TAACTTTTGT TGGAAACAAAT 3640  
 TTTCCAAACC GCCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCGTTTAAAG AGGCTTAGGG 3710  
 ATCACTAAGG GGATTCTAG AAGAGCGACC TGTAATCTTA AGTATTTACA AGACGAGGCT AACCTCCAGC 3780  
 GAGCGTGACA GCGGAGGAG GGTGCGAGGC CTGTTCAAAAT GCTAGTCCA TAAATAAGC AATTTCCTCC 3850  
 GGCAGTTTCT GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTGTAGT AGTTCAAGT TTGCCGACCT 3920  
 CAGCTACAGC ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCCC CCTTAGATCC AAACCTGAGC 3990  
 AACCCGGAGT CTGGATTCTT GGGGAAGTCT CAGCTGTCTT GCGGTTGTGC CCGGGCCCCA GGTCTGGAGG 4060  
 GGACCACTGG CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CCGGCTCTCAG AGCTCTGCAG TCCGAGGCTT 4130  
 GGAGCCAGGT GCCTGGAACC CGAGGCTGCC CTCCACCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCCT 4200  
 CATCTGCCAG ACAGAGTGCC GGGGCCAGG GTCAAGCCCG TTGTGGCTGG TGTGAGGCGC CCGGTGGCGC 4270  
 GTCAGCAGGA GCGCCTGGCT CCATTTCCTC CCCTTCTCG ACAGGACCGC CCGGTGGGTG GATTACAGA 4340  
 TTTGGGGTGG TTTGCTCATG GTGGGGACCC CTCGCCGCTC GAGAACCTGC AAAGAGAAAT GACGGGCTCT 4410  
 TGTCAAGGAG CCAAGTCCG GGGGAAGTGT TGCAGGAGG CACTCCGGGA GGTCCCGCGT GCCCGTCCAG 4480  
 GGAGCAATGC GTCTCGGGT TCGTCCCGAG CCGCGTCTAC GCGCCTCCGT CCTCCCTTC ACGTCCGGCA 4550  
 TCGTGGTGC CCGGAGCCCG ACGCCCCGCG TCCGAGCTGC GAGGCAGCCC TGGGTCTCCG GATCAGGCCA 4620  
 GCGGCCAAAG GGTGCCCGCA CGCACTGTT CCCAGGGCT CCACATCATG GCGCTCCCT CCGGTACCC 4690

Fig. 4 (Fortsetzung)

```
CACAGCCTAG GCCGATTCTGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GCGGGGGAAG CGCGGCCAG ACCCCCGGGT CCGCCCGGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CCACTGGATT CGCGGGCACA GACGCCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCT CCCCTTCCTT TCCGCGGGCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCCC GCGATG 5126
```



Fig. 5

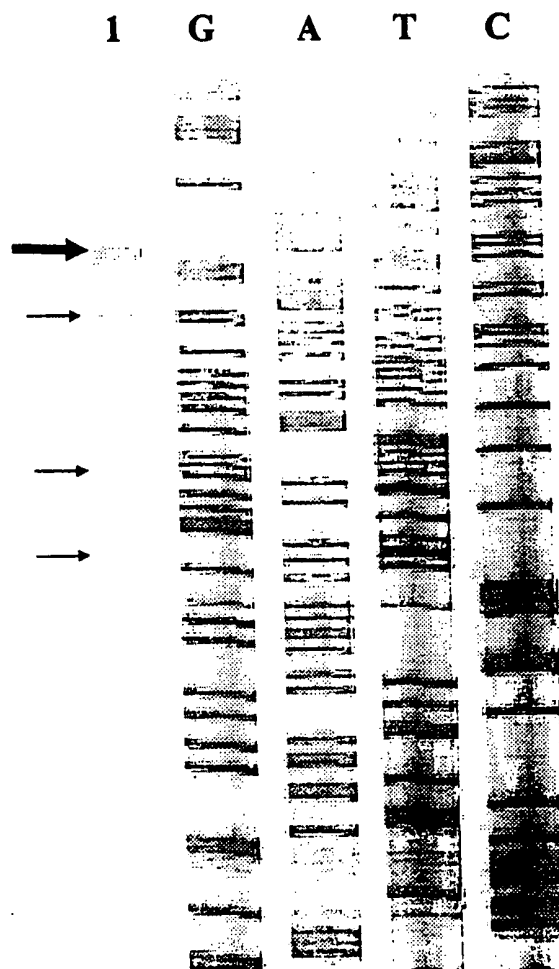
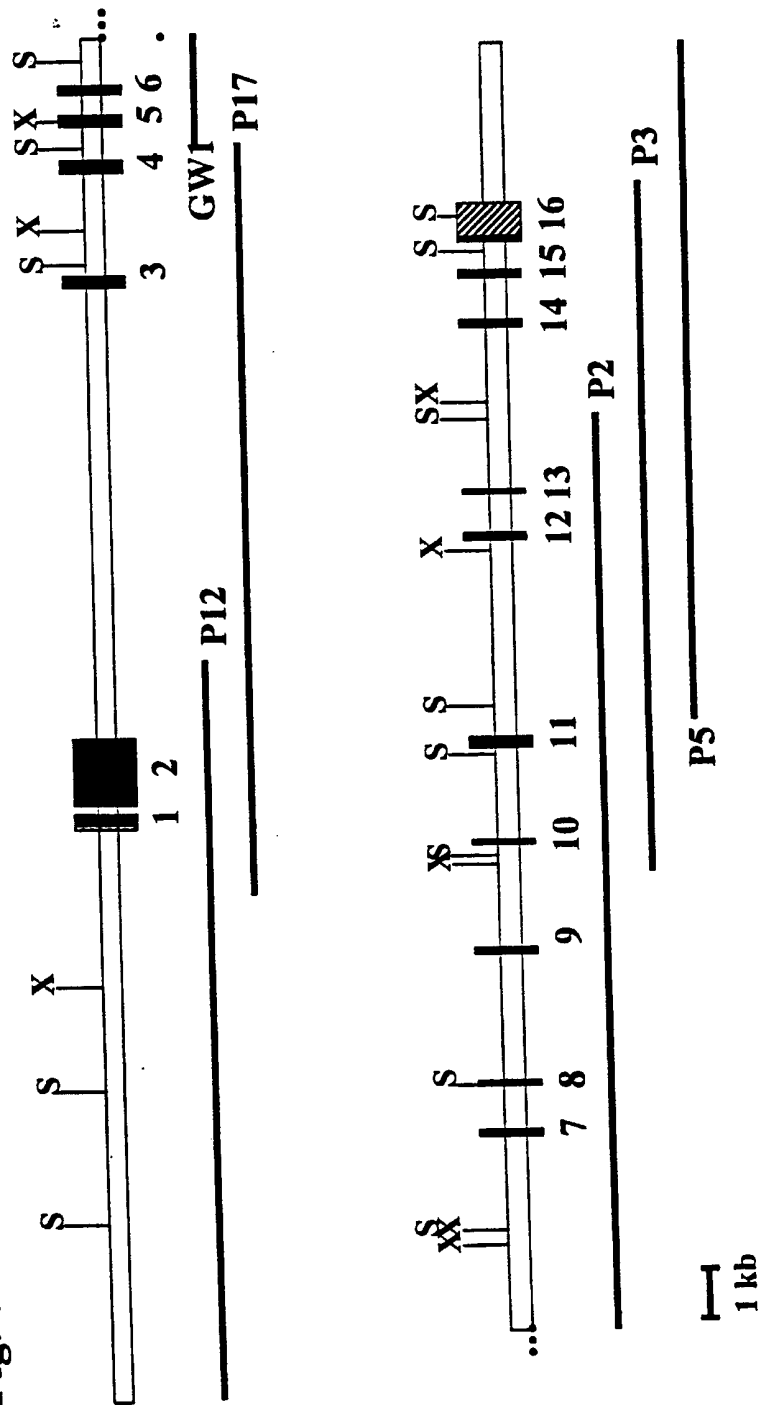


Fig. 6

GTTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCCG GCCACCCCCG CGATGCCCGG 70  
 CGCTCCCCCG TGCCGAGCCG TGCGCTCCCT GCTGCGCAGC CACTACCCCG AGGTGCTGCC GCTGGCCACG 140  
 TTCGTGCGGC GCCTGGGGCC CCAGGGCTGG CGGCTGGTGC AGCGCGGGGA CCCGGCGGCT TTCGCGCGC 210  
 TGGTGGCCCA GTGCTGGTG TGCGTGCCCT GGGACGACAG GCGCGCCCCC GCGCGCCCTT CCTTCCGCCA 280  
 GGTGTCTGTC CTGAAGGAGC TGGTGGCCCG AGTGTGTCAG AGGCTGTGCG AGCGCGGCGC GAAGAAGCTG 350  
 CTGGCCCTTCG GCTTCGCGCT GCTGGACGGG GCGCGCGGGG GCGCCCCCGA GGCCTTCACC ACCAGCGTGC 420  
 GCAGCTACCT GCCAACACG GTGACCGAGC CACTGCGGGG GAGCGGGGCG TGGGGGCTGC TGTGCGCCG 490  
 CGTGGGCGAC GACGTGCTGG TTCACCTGCT GGCACGCTGC GCGCTCTTTG TGTGGTGGC TCCAGCTGC 560  
 GCCTACCAGG TGTGCGGGCC GCGGCTGTAC CAGCTCGGCG CTGCCACTCA GCGCGGCCCC CCGCCACACG 630  
 CTAGTGGACC CCGAAGGCGT CTGGGATGCG AACGGGCTG GAACCATAGC GTCAGGGAGG CCGGGGTCCC 700  
 CCTGGGCGTG CCAGCCCCGG GTGCGAGGAG GCGCGGGGCG AGTGCCAGCC GAAGTCTGCC GTTGCCCAAG 770  
 AGGCCCAGGC GTGGCGCTGC CCCTGAGCCG GAGCGGACGC CCGTTGGGCA GGGGTCTCGG GCCCACCCGG 840  
 CAGGACGCG TGAGCCGAGT GACCGTGGTT TCTGTGTGGT GTCACCTGCC AGACCCGCGC AAGAAGCCAC 910  
 CTCTTTGGAG GGTGCGCTCT CTGGCACGCG CCACTCCAC CCATCCGTGG GCGGCCAGCA CCACGCGGGC 980  
 CCCCACCCA CATCGCGGCC ACCACGTCCC TGGGACACGC CTTGTCCCCC GGTGTACGCC GAGACCAAGC 1050  
 ACTTCCCTTA CTCCTCAGGC GACAAGGAGC AGCTGCGGCC CTCCCTCTTA CTCAGCTCTC TGAGGCCACG 1120  
 CCTGACTGGC GCTCGGAGGC TCGTGAGAC CATCTTCTG GGTTCAGGC CCTGGATGCC AGGGACTCCC 1190  
 CGCAGGTTGC CCGGCTGCC CCAGCGCTAC TGGCAATGC GCGCCCTGTT TCTGGAGCTG CTTGGGAACC 1260  
 ACGGCGATG CCGCTACGGG GTGCTCTCA AGACGCACTG CCGGCTGCGA GCTGCGGTCA CCCCAGCAGC 1330  
 CGGTGTCTGT GCGCGGAGA AGCCCCAGGG CTCTGTGGCG GCGCCGAGG AGGAGGACAC AGACCCCGT 1400  
 CGCTGTGTC AGCTGCTCCG CCAGCAGAGC AGCCCCGGC AGGTGTACGG CTTGTCGCGG GCCTGCCTGC 1470  
 GCGGCTGGT GCGCCAGGC CTCTGGGCT CCAGGCACAA CGAACGCGC TTCCTCAGGA ACACCAAGAA 1540  
 GTTCATCTCC CTGGGGAAGC ATGCCAAGCT CTCGCTGAG GAGCTGACGT GGAAGATGAG CGTGGCGGAC 1610  
 TGGCGTTGGC TGCGCAGGAG CCCAGGGGTT GCGTGTGTT CCGCCGAGA GCACCGTCTG CGTGAGGAGA 1680  
 TCCTGGCCAA GTTCTGCAC TGGCTGATGA GTGTGTACGT CGTGAGCTG CTCAGGTCTT TCTTTATGT 1750  
 CACGGAGACC ACGTTTCAA AGAACAGGCT CTTTCTTAC CGGAAGAGTG TCTGGAGCAA GTTGCAAGC 1820  
 ATTGAATCA GACAGCACTT GAAGAGGGTG CAGCTGCGGG AGCTGTGCGA AGCAGAGGTC AGGCAGCATC 1890  
 GGGAAAGCAG GCGCGCCCTG CTGACGTCCA GACTCCGCTT CATCCCCAAG CCTGACGGGC TGGCGCCGAT 1960  
 TGTGAACATG GACTACGTCG TGGAGCCAG AACGTTCCGC AGAGAAAGA GGGCCGAGCG TCTCAGCTG 2030  
 AGGGTGAAG CACTGTTTCA CGTCTCAAC TACGAGCGGG GCGCGCGCCC GCGGCTCCTG GCGGCTCTG 2100  
 TGCTGGGCTT GAGCAGATATC CACAGGGGCT GCGCACCTT CGTGTGCGT GTGCGGGCCC AGGACCCGCC 2170  
 GCCTGAGCTG TACTTTGTCA AGGTGGATG GACGGGCGCG TACGACACCA TCCCCCAGGA CAGGCTCAGC 2240  
 GAGGTATCG CCAGCATCAT CAAACCCAG AACACGTACT GCGTGCCTG GTATGCCGTG GTCCAGAAGG 2310  
 CCGCCCATGG GCACGTCCGC AAGGCCITCA AGAGCCACGT CTCTACCTTG ACAGACCTCC AGCCGTACAT 2380  
 GCGACAGTTC GTGGCTCACC TGCAGGAGC CAGCCGCTG AGGGATGCGG TCGTCATCGA CAGAGCTCC 2450  
 TCCTGAATG AGGCCAGCAG TGGCTCTTC GACGCTCTTC TACGCTTCAT GTGCCACCAC GCGGTGCGCA 2520  
 TCAGGGGCAA GTCCTAGTC CAGTGCCAGG GGATCCCGCA GGGCTCCATC CTCTCCACGC TGTCTGTCAG 2590  
 CTTGTGCTAC GCGCAGATGG AGAACAGCT GTTTCGGGG ATTCGGGCGG ACGGGCTGCT CCGGCTTTG 2660  
 GTGGATGATT TCTTGTGGT GACACCTCAC CTCACCCAGC CGAAACCTT CCTCAGGACC CTGGTCCGAG 2730  
 GTGTCCCTGA GTATGGCTGC GTGGTGAAT TCGGAAGAC AGTGGTGAAC TTCCCTGTAG AAGACGAGGC 2800  
 CCTGGGTGGC ACGGCTTTTG TTCAGATGCC GCGCCACGGC CTATTCCTCT GGTGCGGCTT CTTGCTGGAT 2870  
 ACCCGAGACC TGGAGGTGCA GAGCGACTAC TCCAGCTATG CCGGACCTC CATCAGAGCC AGTCTCACCT 2940  
 TCAACCGCG CTTCAGGCT GGGAGGAACA TCGCTCGCAA ACTCTTTGG GTCTTGCGGC TGAAGTGTCA 3010  
 CAGCCTGTTT CTGGATTTC AGGTGAACAG CCTCCAGAG GTGTGCACCA ACATCTACAA GATCCTCTG 3080  
 CTGAGGCGT ACAGGTTTCA CGCATGTGTG CTGAGCTCC CATTTATCA GCAAGTTTGG AAGAACCCCA 3150  
 CATTTTCTT GCGGCTCATC TCTGACACGG CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACGAGG 3220  
 GATGTGCTG GGGGCCAAGG GCGCGCGCGG CCTCTGCCC TCCGAGGCGG TGCAGTGGCT GTGCCACCA 3290  
 GCATTCTGTC TCAAGCTGAC TCGACACCGT GTCACCTACG TGCCACTCCT GGGGTCACTC AGGACAGCCC 3360  
 AGACGAGCT GAGTCGGAAG CTCCCGGGGA CGACGCTGAC TGCCCTGGAG GCGCGAGCCA ACCCGGCACT 3430  
 GCCCTCAGAC TTCAAGACCA TCCTGAGTGC ATGGCCACCC GCGCCACAGC AGGCGGAGAG CAGACACAG 3500  
 CAGCCCTGTC ACGCCGGGCT CTACGTCCCA GGGAGGGAGG GCGCGCCAC ACCCAGGCC GCACCGCTGG 3570  
 GAGTCTGAG CTTGAGTGAG TGTTTGGCG AGGCTGCGAT GTCCGGCTGA AGGCTGAGTG TCCGGCTGAG 3640  
 GCCTGAGCGA GTGTCCAGCC AAGGGCTGAG TGTCCAGCAC ACCTGCCGTC TTCACTTCCC CACAGGCTGG 3710  
 CGCTCGGCTC CACCCAGGG CCAGCTTTTC CTCACAGGA GCGCGGCTTC CACTCCCCAC ATAGGAATAG 3780  
 TCCATCCCCA GATTCCGCAT TGTTCACCCC TCGCCCTGCC CTCCCTTGCC TTCCACCCCC ACCATCCAGG 3850  
 TGGAGACCT GAGAAGGACC CTGGGAGCTC TGGGAATTG GAGTGACCAA AGGTGTGCCC TGTACACAGG 3920  
 CGAGGACCT GCACCTGGAT GGGGTCCTT GTGGGTCAA TGGGGGGAG GTGCTGTGG AGTAAATAC 3990  
 TGAATATATG AGTTTTTCAG TTTTGAAGAA AAAAAAAAAA AAAAAAAAAA AA 4042

Fig. 7



**Fig. 8A**

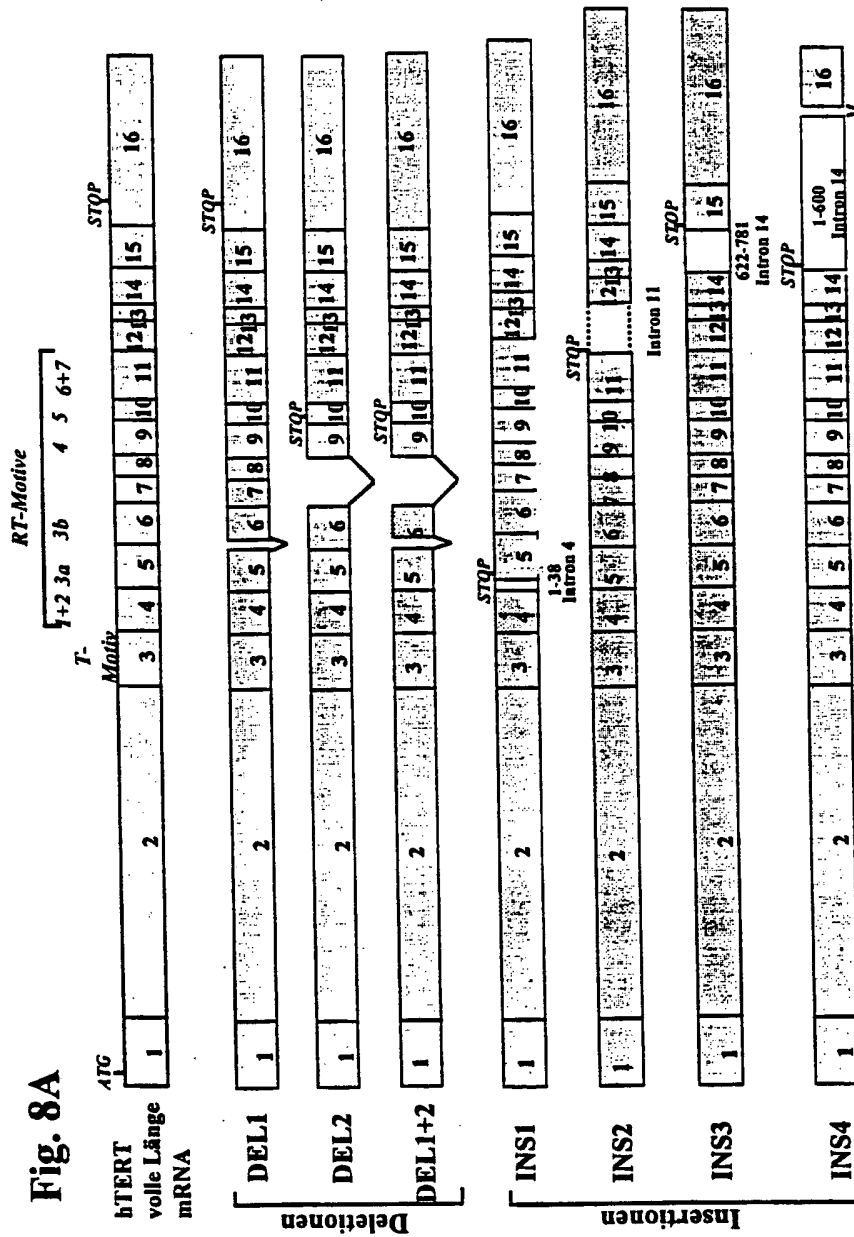
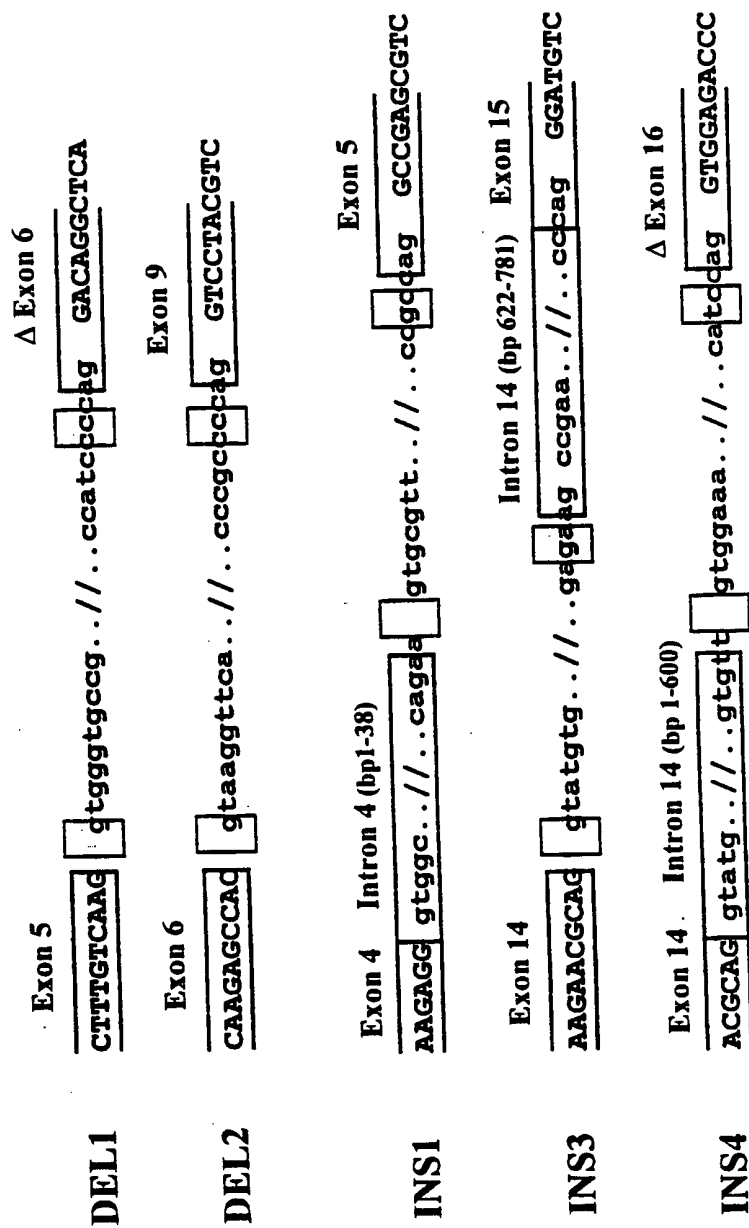


Fig. 8B



**Fig. 9**

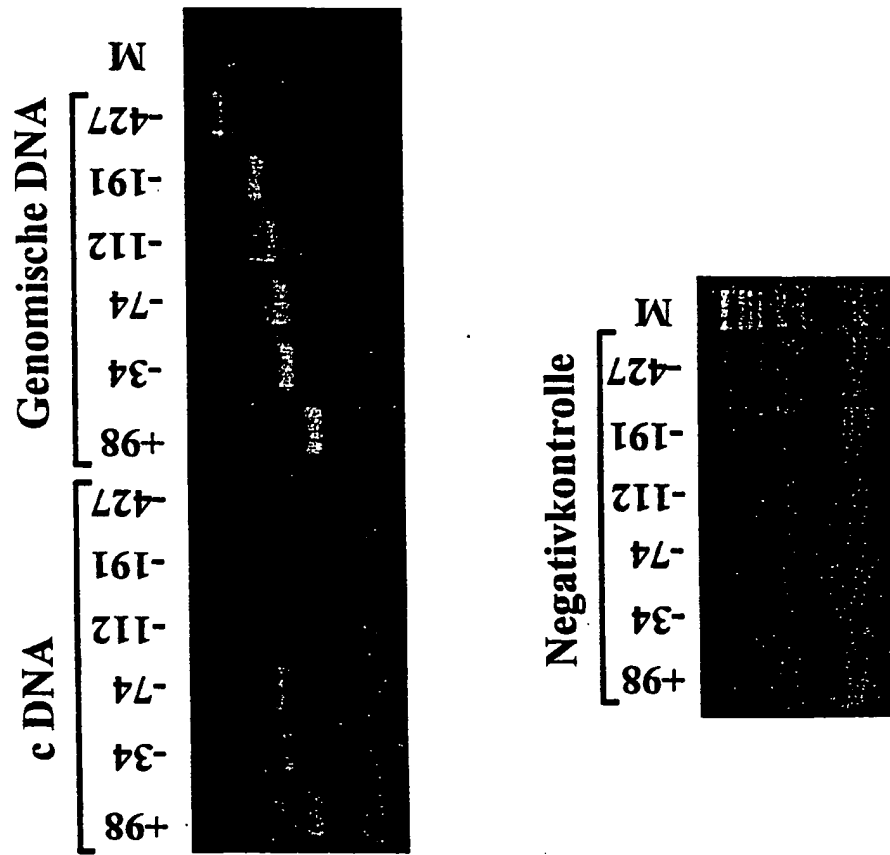


Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204  
 ATGAGACCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAGCA TCTTCTCTGG CCACAGTGGA -11134  
 AAAAAACCAG AATCAACAA CAAGAGGAAT TTTGAAAAC ATACAAACAC ATGAAAATTA AACAAATATAC -11064  
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTA AAAAGGA AATTGAAAAA TTTATTTAAG CAAATGATAA -10994  
 CGGAAACATA ACCTCTCAA ACCCACCAGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924  
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAG -10854  
 GCCAAGGCGG GCAGATCGCC TGAGGTGAGG AGTTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTCG -10784  
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG -10714  
 GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGGCCCATTT GACTCCAGCC -10644  
 TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT -10574  
 GATGCACCTT AAAGAAGTAG AAAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAAGAAA AGAATAATA -10504  
 AAGATCAGAG CAGAAATAAA TGAAACTGAA AGATAACAAT ACAAAGATC AACAAAATTA AAAGTTGGTT -10434  
 TTTTAAAAA ATAAACAAAA TTGACAAACC TTTGCCAGA CTAAGAAAAA AGGAAAGAAG ACCTAAATAA -10364  
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAATT CAAAGGATCA CTAGAGGCTA -10294  
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCTAGA TGCATACAAC -10224  
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACAC GACCAATAAC AATAATGGGA TTAAGCCAT -10154  
 AATAAAAGT CTCTAGCAA AGAGAAGCCC AGGACCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT -10084  
 AAAGAAGAT GAATTCAT CTACTCAA CTATTCTGAA AATAGAGGA AAGAATACTT CCAACTCAT -10014  
 TCTACATGGC CAGTATTACC CTGATTCCAA AACCCAGCAA AACACATCA AAAACAAACA AACAAAAAA -9944  
 CAGAAAGAAA GAAACTACA GGCCAAATC CTGTATGAAT ACTGATACAA AATCCTCAA CAAACACTA -9874  
 GCAAAACAAA TTAACAACA CCTTCGAAAG ATCATTCT GTGATCAAGT GGGATTTATT CCAGGGATGG -9804  
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAACATA -9734  
 TATGATTATT TCACCTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTATGATA AAAACCTCA -9664  
 AAAAAACCAGG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGGCATCCCA GCACCTGGG -9594  
 AGGCCAAGGT GGGATGATT CTGGGCCCCA GGAGTTTGG ACTAGCCTGG GCAACAAAT GAGACCTGGT -9524  
 CTACAAAAA CTTTTTAA AAATTAGCCA GGCATGATG CATATGCCTG TAGTCCAGC TAGTCTGGAG -9454  
 GGTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCACTGAGC CATGAACATG TCACTGTACT -9384  
 CCAGCCTAGA CAACAGAACA AGACCCCACT GAATAAGAG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG -9314  
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244  
 AAGAAACATA TTTCAACATA ATAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGTGA -9174  
 AGCCTTCTCT CTAAGATCTG GAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA -9104  
 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAA -9034  
 TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGGAAA GACTTAAGAC ACCACTAAA AACTATTAGA -8964  
 CTGAAATTT GGTACAGCAG GATACAAAT CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC -8894  
 AAACAATCTG AAAAAGAAAC CAAAAAGCA GCTACAAATA AAATTAACA GCTAGGAATT AACCAAAGAA -8824  
 GTGAAAGAT TCTACAATGA AAATATAAA ATGTTGATAA AAGAAATGA AGAGGGCACA AAAAAGAAA -8754  
 AGATATTCCA TGTTTATAGA TTGGAAGAAT AAATACTGTT AAAATGTCCA TACTACCCAA AGCAATTTAC -8684  
 AAATTCATG CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT -8614  
 TTGTACAGAA CCACAAAAGA CCCAGATAG CCAAAGCTAT CCTGACCAA AAGAACAAA CTGGAAGCAT -8544  
 CACATTACCT GACTTCAAAT TATACTACAA AGCTATAGTA ACCCAAATA CATGGTACTG GCATAAAAAC -8474  
 AGATGAGACA TGGACCAGAG GAACAGAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAACCTA -8404  
 TTTTTCAGAA AGGTGCCAAG AACATACTTT GGGGAAAAGA TAATCTCTT AATAAATGGT GCTGGAGGAA -8334  
 CTGGATATCC ATATGCAAAA TAACAATACT AGAAGCTCTG CTCTCACCAT ATACAAAAGC AAATCAAAT -8264  
 GGATGAAAGG CTTAAATCTA AAACCTCAA CTTTGCACT ACTAAAAGAA AACACCGGAG AAATCTCTCA -8194  
 GGACATTGGA GTGGGCAAG ACTTCTTGAG TAATCCCTG CAGGCACAGG CAACCAAAGC AAAACAGAC -8124  
 AAATGGGATC ATATCAAGTT AAAAAGCTT TGCCAGCAA AGGAAACAAT CAACAAAGAG AAGAGACAAC -8054  
 CCACAGAATG GGAGAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA -7984  
 GCTCAAACCTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAAT AAGCAAAAGA TCTGGGTAGA -7914  
 CATTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAATGTG CTCAACACCA CTGATCATCA -7844  
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAGAC -7774  
 AGGCAATAAC AATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAA -7704  
 TTGCTACCAC TATGGGAAC AGTTTGAAAG TTCCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC -7634  
 CCATTGCTAG GTATATACTC CAAAAAAGG AATCAGTGTA TCAACAAGCT ATCTCCACTC CCACATTTAC -7564  
 TGCAGCACTG TTCATAGCAG CCAAGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG -7494  
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA -7424  
 CAGCATGGGG GGCACCTGGT AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAC TTTTCATGTT -7354  
 CTCCCTTACT TGTGGGAGCA AAAATTA AAA CAATTGACAT AGAATAGAG GAGAATGGTG GTTCTAGAGG -7284  
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA AGAGTATAAT -7214  
 TGGGTTGTT GTACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCCATTTAC CCTGATGTGA -7144  
 TTATTACACA TTGTATGCCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCCTA CTATATTA -7074  
 AATTAATAAT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCAGC ACTTTGGGAG GCCGAGGCGG -7004  
 GTGGATCACC TGAGGTCAGG AGTTTGAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTACTAAAGA -6934  
 TACAAAAAT AGCCAGCGT GGTGGGCAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA -6864  
 TTGCTTGAAC CTGGGAGCG GAGGTTGAG TGAGCCGAGA TCATGCCACT GCACTGCAGC CTGGGTGACA -6794  
 GAGCAAGACT CCATCTCAA AAAAAACAA AAAAAAGAA ATTAAAAATG TAATTTTAT GTACCGTATA -6724  
 AATATATACT CTACTATATT AGAAGTTAAA AATTAAAAA ATATAAAG GTAATTAACC ACTTAATCTA -6654  
 AAATAAGAAC AATGTATGTG GGGTTTCTAG CTTCTGAAGA AGTAAAGTT ATGGCCACGA TGGCAGAAAT -6584

Fig. 10

GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC -6514  
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAGG -6444  
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374  
 CACCGTCTCT TCATTACCGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304  
 GTTAAACTTA ATCTGTATGA ATCCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAAGAA TTAGAGTACC -6234  
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164  
 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094  
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGGAAG -6024  
 AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954  
 CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACTTGA -5884  
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCCTTCT TAAACAGAA AGTCATGGAA -5814  
 GCACCTTTCT CAAGGGA AAAACAGGCA CCAGACGCCC GCTCTGCGGT CATTTACCTC TTTCTCTCTCT CCCTCTCTTG -5744  
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCCTGGGA GCTTCTCCGA GCCCGTGTG AGGACCCCTG -5674  
 TGCAAGGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604  
 TGGCTGGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCATGAG TAAATCAAC CTTTCCACAT -5534  
 CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTAAATT TCATCAATA ACATTCAGGA CTCGCAAAAT -5464  
 CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394  
 ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAG GTACACGAGG -5324  
 AGAGGCTTGG GCGGCAGGGC TATGAGCAGC GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254  
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGCGGC CACCGTGCCT -5184  
 GTGACTCAGG ACCCATACCC GGCTTCCTGG GCCCACCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114  
 CCGTGGAAC GAACATGACC CTTGCCCTGC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTGCA -5044  
 GGAATGGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATATT CATCTTGACC -4974  
 CCAAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAACTCAG TACAAACACC -4904  
 ACTCTTTTAC TAGGCCACCA GAGCAGGSCC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834  
 GCTTTAGACC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTC TCTAGACTAG TAGACCCTGG -4764  
 CAGGCACTCC CCCAGATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694  
 CTGGGGTGCC AACTGAGGC CAGCCCTGTC TCCACACCCT CCGCCTCCAG GCCTCAGCTT CTCCAGCAGC -4624  
 TTCTTAAAC CTGGGTGGGC CGTGTCCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGC -4554  
 ACGTAGCTCG CACGGTTCCT CCTCACATGG GGTGTCTGTC TCCTTCCCA ACACCTACAT CGCTTGAAG -4484  
 GAGGAGATTG TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGAGGTT -4414  
 CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCGTCTC CTGTCTATG CCGGGGCTG -4344  
 CCGGTGTGTT CTCTGTGTTT TGTGCTCCTT TCCAGCTCCA GCTGCGTGTG TCTCTGCGCC CTAGGCTCTC -4274  
 GGGGTTTTTA TAGGCATAGG ACGGGGCGT GGTGGGCCAG GCGCTCTTG GGAATGCAA CATTTGGGTG -4204  
 TGAAAGTAGG AGTGCTGTG CTCACCTAGG TCCACGGGCA CAGGCTGGG GATGGAGCCC CCGCCAGGGA -4134  
 CCCGCCCTTC TCTGCCCAGC ACTTTCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TGCACAAGC -4064  
 ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGACATTT GCCCCACAGC CCTGGGAATT -3994

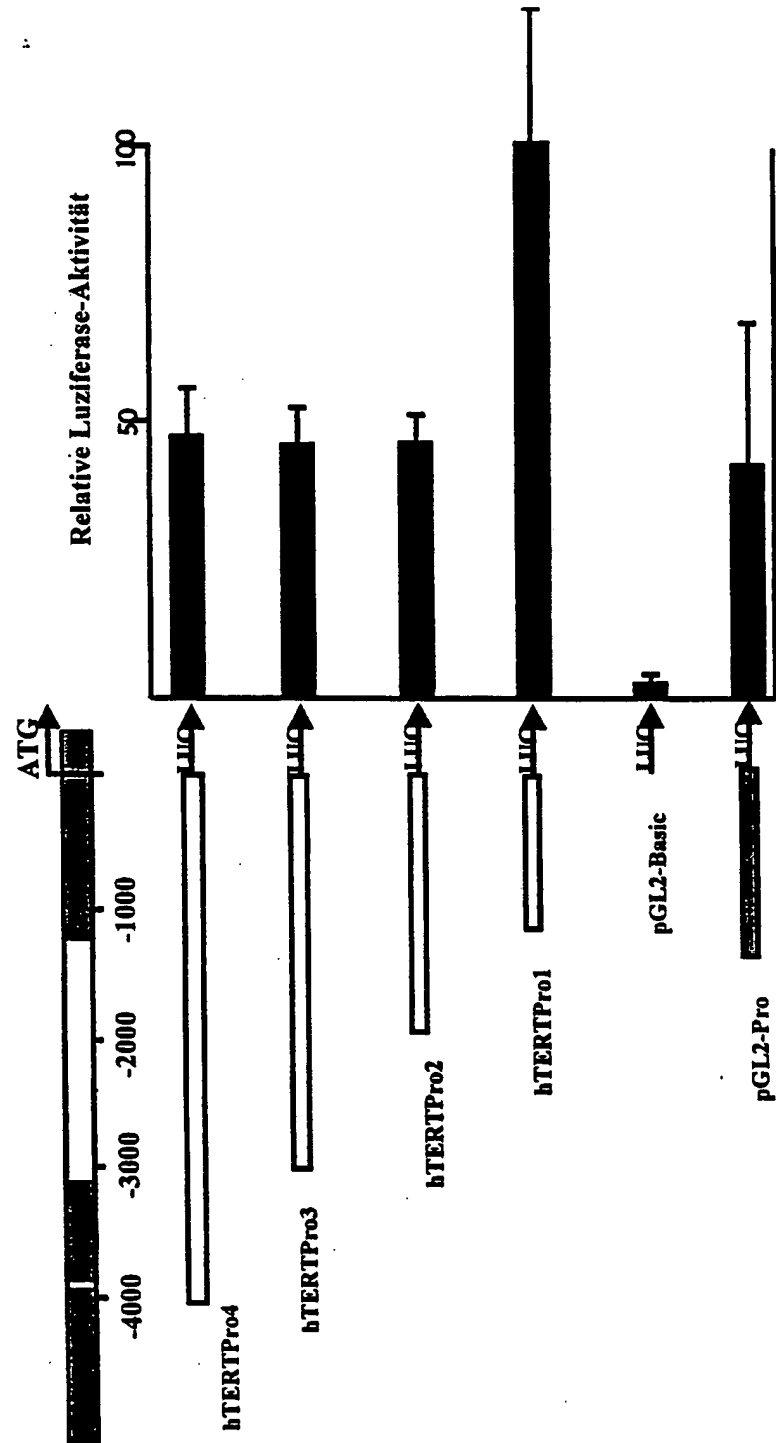
c-Myc  
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTTATT TTAATAGCTA -3924  
 CAAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854  
 GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784  
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAAG AAAGAATTTC ACCCCATGGC -3714  
 GGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACAGC CCGTTTACTA -3644  
 AAGCCAGTTT CTGTTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574  
 GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CTTGGGCAGG ATAATGCTCT AGAGATGCC -3504  
 ACGTCTGAT TCCCCAAAC CTGTGGACAG AACCCGCGG CCCCCAGGGC CTTTGCAGGT GTGATCTCCG -3434  
 TGAGGACCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCGAAAA GTAATCCAGG GGTCTGGGA -3364  
 AGAGGCGGGG AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAG -3294  
 GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCTCCAG AAGCTGGAAA AAGCGGGGAA GGGACCTCC -3224  
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC TCCGGCTCC -3154  
 GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGG ATTTGCAGAA GCAACAGGAA -3084  
 ACCCATGCAC TGTGAATCTA GGATTATTC AAAACAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA -3014  
 GTGCCTCCGG GCAAGGGCAG GGCAGGCAG AGTGATTTTA TTTAGCTATT TTATTTTATT TACTTACTTT -2944  
 CTGAGACAGA GTTATGCTCT TGTGCCCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874  
 CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTCAGGC GTGCACCACC -2804  
 ACACCCGGCT AATTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAGCTGA TCTCAAAATC -2734  
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664  
 GGCCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594  
 AATTTCCCT TTAAGCAGGA GTTACCCTCC TTTGATATTT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524  
 CGTCTCTTGA CATATTCACA GTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGGCAGGTGG -2454  
 GAGGCTGCAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGGCG -2384  
 AAGTGTGGAC ACTGTCCGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314  
 CTCTACTCT ACTGGGATTG AGCCCCCTCC CTATCCCCC CCAGGGGCGAG AGGAGTTCCT CTCACTCCTG -2244  
 TGGAGGAAGG AATGATACTT TGTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTGGTTTG -2174  
 TTTGTTTTGT TTTGAGAGGC GGTTCCTCA TTGTTGCTCA GGCTGGAGGG AGTGCAATTC CCGCATCTG -2104  
 GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTCCGC CTCCCATTTG GCTGGGATTA -2034  
 CAGGCACCCG CCACCATGCC CAGCTAATTT TTTGATTTT TAGTAGAGAC GGGGGTGGT GGGGTTTACC -1964



ATGTTGGCCA GGCTGGTCTG GAACCTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894  
GATTACAGGT GTGAGCCACC ATGCCCAGCT CAGAATTAC TCTGTTTAGA AACATCTGGG TCTGAGGTAG -1824  
GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAACAAT-Box  
GTTAAGCCAA TSATAGAATT TTTTATTGT TGTTAGAACA -1754  
CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA -1684  
ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614  
GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTCCCT CTTTAAAT TGTGTTTTCT -1544  
ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAACT TAACTTTTGT TGAACAAAT TTTCCAACC -1474  
Spi  
GCCCTTTTG CCTAGTGGCA GAGACAATC ACAACACAG CCCTTTAAAG AGGCTTAGGG ATCACTAAGG -1404  
GGATTCTAG AAGAGCGACC TGTAATCTTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334  
GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTTCCTCC GGCAGTTTCT -1264  
GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTAGC ATTTCACTGT TTGCCGACCT CAGCTACAGC -1194  
ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCCC CTTAGATCC AAACCTGAGC AACCCGGAGT -1124  
CTGGATTCTT GGGAACTCCT CAGCTGTCCT GCGGTTGTGC CGGGGCCCA GGTCTGGAGG GGACCACTGG -1054  
CCGTGTGGCT TCTACTGCTG GGCTGGAACT CGGGCCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984  
GCCTGGACCC CGAGGCTGCC CTCACCCCTG TCGGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG -914  
ACAGAGTGCC GGGGCCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CGGTGCGCG GCCAGCAGGA -844  
GCGCCTGGCT CCATTCCCA CCTTTCTCG ACGGGACGC CCCGTGGGT GATTAACAGA TTTGGGGTGG -774  
TTTGCTCATG GTGGGGACCC CTCGCCCCT GAGAACCTGC AAAGAGAAAT GACGGGCCCTG TGTCAAGGAG -704  
CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCT GCGCGTCCAG GGAGCAATGC -634  
GTCTCGGGT TCCTCCGAG CCGCTCTAC GCGCCTCCGT CCTCCCTTC ACCTCCGCA TTCGTGGTGC -564  
CCGGAGCCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494  
GGTCGCCGA CGCACTGTT CCCAGGCCT CCACATCATG GCCCCTCCCT CGGGTTACCC CACAGCCTAG -424  
GCCGATTCTA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CTGCACCTT GGGAGCGCGA GCGGCGCG -354  
GGCGGGGAAG CGCGGCCAG ACCCCCGGT CGGCCGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC -284  
CCAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC CGAGTGCCG GAGGGAAGTGG GGACCCGGGC -214  
ACCCGTCCTG CCCCTTACC TTCCAGTCC GCCTCCTCG CGCGGACCC GCCCTCC GACCCCTCC -144  
GGGTCCCCG CCCAGCCCC TCCGGGCCCT CCCAGCCCT CCCCCTCCTT TCCGCGGCC CGCCCTCTCC -74  
TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGTGCG ACGTGCGAAG CCCTGGCCCC GGCCACCCCC -4  
GCGATG

15 / 15

Fig.: 11



## SEQUENZPROTOKOLL

&lt;110&gt; Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der Si-Region vom Gen  
der humanen katalytischen Telomerase-Untereinheit und  
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

&lt;140&gt;

&lt;141&gt;

15 <160> 20

&lt;170&gt; PatentIn Vers. 2.0

&lt;210&gt; 1

20 <211> 5126

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 1

25 gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60  
gggtaatgaa gtgggtgtgca ggaaatggcc atgtaaaatta cacgactctg ctgatgggga 120  
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180  
tcgggtgtga caagccatga caaaactcag tacaacacacc actcttttac taggcccaca 240  
gagcacgggc cacacccctg atatatgaag agtccaggag agatgaggct gctttcagcc 300  
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagacctgg 360  
30 caggcactcc cccaaattct agggcctggg tgctgtctcc cgaggggcgc atctgcctcg 420  
gagactcagc ctgggggtgcc acactgaggc cagccctgtc tccacaccct ccgcctccag 480  
gcctcagctt ctccagcagc ttcctaaacc ctgggtgggc cgtgttccag cgctactgtc 540  
tcacctgtcc cactgtgtct tgtctcagcg acgtagctcg cacggttcct cctcacatgg 600  
ggtgtctgtc tccttcccca acactcacat gcgttgaagg gaggagattc tgcgcctccc 660  
35 agactggctc ctctgagcct gaacctggct cgtggccccc gatgcagggt cctggcgctc 720  
ggctgcacgc tgacctccat ttccaggcgc tcccgcgtct cgtcatctg ccggggcctg 780  
ccgggtgtgt ctctctgttc tgtgtctcct tccacgtcca gctgcgtgtg tctctgccc 840  
ctagggtctc ggggttttta taggcatagg acggggggcgt ggtggggcag ggcgctcttg 900  
40 ggaaatgcaa catttgggtg tgaaagttag agtgccctgt ctacactagg tccacgggca 960  
caggcctggg gatggagccc ccgccaggga cccgcccttc tctgccagc acttctctgc 1020  
ccccctccct ctggaacaca gagtggcagt ttccacaagc actaagcacc ctcttcccaa 1080  
aagaccagc attggcaccc ctggacattt gccccacagc cctgggaatt cactgacta 1140  
cgcacatcat gtacacactc ccgtccacga ccgacccccc ctgttttatt ttaatagcta 1200  
caaagcaggg aaatccctgc taaaatgtcc tttaacaaac tggttaaaca aacgggtcca 1260  
45 tccgcacggg ggacagtccc tcacagtga gaggaacatg ccgtttataa agcctgcagg 1320  
catctcaagg gaattacgct gagtcaaaac tgcacacctc atgggatagc tacgcaacat 1380  
gctcaaaaag aaagaatttc accccatggc aggggagtggt ttaggggggt taaggacggg 1440  
ggggggcgga gctgggggct actgcacgca ccttttacta aagccagttt cctgggtctg 1500  
atgggtattg ctcatgtatg ggagactaac cataggggag tggggatggg ggaacccgga 1560  
50 ggctgtgcca tctttgccat gcccgagtgt cctgggcagg ataagtctct agagatgccc 1620  
acgtcctgat tcccccaaac ctgtggacag aaccgcggcg gcccaggggc ctttgagggt 1680  
gtgatctccg tgaggaccct gaggtctggg atccttcggg actacctgca ggcccgaaaa 1740  
gtaatccagg ggttctggga agaggcgggc aggggggtca gaggggggca gcctcaggac 1800  
gatggaggca gtcagtctga ggctgaaaag ggaggggagg cctcgagccc aggcctgcaa 1860  
55 gcgcctccag aagctggaaa aagcggggaa gggaccctcc acggagcctg cagcaggaa 1920  
gcacggctgg cccttagccc accaggggcc atcgtggacc tccggcctcc gtgccatagg 1980  
agggcactcg cgtgcccctt ctgacatgaa gtgtgtgggg atttgcagaa gcaacaggaa 2040  
acccatgcac tgtgaatcta ggattatttc aaacaaaagg ttacagaaa catccaagga 2100  
cagggtgtaa gtgcctccgg gcaagggcag ggcaggcacg agtgatttta tttagctatt 2160  
60 ttattttatt tacttacttt ctgagacaga gttatgctct tgggtcccag gctggagtgc 2220  
agcggcatga tcttggctca ctgcaacctc cgtctcctgg gttcaagcaa ttctcgtgcc 2280  
tcagcctccc aagtagctgg gatttcaggc gtgcaccacc acaccggcct aattttgtat 2340  
ttttagtaga gatgggcttt caccatgttg gtcaaactga tctcaaaatc ctgacctcag 2400  
gtgatccgcc cacctcagcc tcccaaatg ctgggattac aggcattgag cactgcacct 2460

2 / 18

5 ggccctattta accatttttaa aacttccctg ggctcaagtc acaccactg gtaaggagtt 2520  
 catggagttc aatttccctt tttactcagga gttaccctcc tttgataatt tctgtaattc 2580  
 ttcgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640  
 tccccggga cccactgcag gggcagctgg gaggtgcag gcttcaggte ccagtggggt 2700  
 tgccatctgc cagtagaaac ctgatgtaga atcagggcgc aagtgtggac actgtcctga 2760  
 atctcaatgt ctcagtgtgt gctgaacat gtagaaatta aagtcacatcc ctccactctc 2820  
 actgggattg agccccctcc ctatcccccc ccaggggcag agggagtctct ctccactcctg 2880  
 tggaggaaag aatgatactt tgttattttt cactgctggt actgaatcca ctgtttcatt 2940  
 10 tttgtgtttg tttgtttttg tttgagaggc ggtttcactc ttgttgctca ggctggaggg 3000  
 agtcgaatgg cgcgactctg gcttactgca gcctctgcct ccagagttca agtgattctc 3060  
 ctgcttccgc ctcccatttg gctgggatta caggcacccg ccaccatgcc cagctaattt 3120  
 tttgtatttt tagtagagac gggggtgggt ggggttcacc atgttgggcca ggctgtctc 3180  
 gaacttctga cctcagatga tccacctgcc tctgcctcct aaagtgtcgg gattacaggt 3240  
 gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300  
 15 gaagctcacc ccactcaagt gttgtggtgt tttacactgt gatgactaag acatcatcag cttttcaaag 3420  
 tgttagaaca ctcttgatgt tttacactgt gatgactaag acatcatcag cttttcaaag 3480  
 acacactaac tgcaccata atactggggt gtcttctggg tatcagcaat ctccattgaa 3540  
 tgccgggagg cgtttcctcg ccactgcacat ggtgttaatt actccagcat aatcttctgc 3600  
 ttcctattct tctcttccct ctcttctcct tttttaaagt atgttggtct ctctgcagag 3660  
 20 aaccagtgtg agctacaact taacttttgt tggaaacaaat tttccaaacc gcccccttgc 3720  
 cctagtggca gagacaattc acaaacacag cccctttaaag aggccttaggg atcactaagg 3780  
 ggatttctag aagagcgacc tgtaactcta agtatttaca agacgagggt aacctccagc 3840  
 gagcgtgaca gcccaggagg ggtgcgaggc ctgttcaaat gctagctcca taaataaagc 3900  
 aatttctctc ggcagtttct gaaagtagga aaggttatat ttaaggttgc gtttgttagc 3960  
 25 atttcagtggt ttgcccagct cagctacagc atccctgcaa ggcctcggga gaccagaag 4020  
 tttctcggcc ccttagatcc aaacttgagc aaccggaggt ctggattcct gggaaagtct 4080  
 cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggtc 4140  
 tctactgctg ggtcggaggt cgggctcctc agctctgcag tccgaggctt ggagccaggt 4200  
 30 gcttgagccc cgaggctgcc ctccaccctg tgcgggcggg atgtgaccag atgttggtc 4260  
 catctcgccg acagagtgcc ggggcccagg gtcaaggcgg ttgtggctgg tgtgaggcgc 4320  
 ccggtgcgcg gccagcagga gcgcctgggt ccatttccca cccttctcgc acgggaccgc 4380  
 cccggtgggt gattaaacaga tttgggggtg tttgctcatg gtggggaccg ctgcgcgctc 4440  
 gagaacctgc aaagagaaat gacgggcctg tgtcaaggag cccaagtcgc ggggaagtgt 4500  
 35 tgcaggaggc caactccggga ggtcccgctg gcccgctcag ggagcaatgc gtccctcgggt 4560  
 tctgtccccc cgcgctctac gcgcctcctc cctcccttcc acgtccggca ttcgtggtgc 4620  
 ccggagcccc acgccccgcg tccggacctg gaggcagccc tgggtctcgc gatcaggcca 4680  
 gcggccaaag ggtcgccgca cgcacctgtt cccagggcct ccacatcatg gcccctcctc 4740  
 cggtgttacc cacagcctag gccgattcga cctctctcgc ctggggccct cgctggcgctc 4800  
 40 cctgcacctt gggagcgcga gcggcgcgcg ggcgggggag cgcggccag acccccgggt 4860  
 ccgcccggag cagctgcgct gtccggggcca ggccgggctc ccagtggatt cgcggggcaca 4920  
 gacgcccagg accgcgctcc ccacgtggcg gagggactgg ggaccggggc acccgctctc 4980  
 ccccttccac ttccagctcc gcctcctcgc cgcggacccc gccccgctcc gaccctctcc 5040  
 ggggtccccg cccagccccc tccgggcccct cccagcccct ccccttctct tccgcccgc 5100  
 45 cgccctctcc tgcggcgcg agttttaggc agcgtgtcgt cctgctgcgc acgtgggag 5126  
 ccctggcccc ggccaccccc gcgatg

&lt;210&gt; 2

&lt;211&gt; 4042

&lt;212&gt; DNA

50 &lt;213&gt; Homo sapiens

&lt;400&gt; 2

55 gtttcaggca gcgctgcgtc ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60  
 cgatgccgcy cgtcccccgc tgcgagcgcc tgcgctccct gctgcgcagc cactaccgcy 120  
 aggtgctgcc gctggccacg ttcgtgcggc gcctggggcc ccagggtctg cggtcgggtc 180  
 agcgcgggga cccggcggtc ttccgcgcgc tgggtggcca gtgcctggtg tgcgtgcctc 240  
 gggacgcagc gccgcccccc gccgccccct ccttccgcca ggtgtcctgc ctgaaggagc 300  
 tgggtggccc agtgctgcag aggtgtgtgc agcgcggcgc gaagaacgtg ctggccttcg 360  
 60 gcttcgcgct gctggacggg gcccgcgggg gcccccccca ggccttcacc accagcgtgc 420  
 gcagctacct gcccacacg gtgaccgagc cactgcgggg gagcggggcg tgggggctgc 480  
 tgcctgcgcy cgtggcgagc gacgtgctgg ttacctgct ggacgctgc gcgctctttg 540  
 tgctggtggc tcccagctgc gcctaccagg tgtgcgggccc gccgctgtac cagctcggcg 600  
 ctgccactca gggccggccc ccgccacacg ctagtggacc ccgaaggcgt ctgggatgcg 660  
 aacgggctcg gaaccatagc gtcaggagag ccgggggtccc cctgggcccgt ccagcccccg 720  
 65 gtgcgaggag gcgcgggggc agtgccagcc gaagtctgcc gttgcccaag agggccaggc 780

3 / 18

gtggcgctgc ccctgagcgg gagcggagcgc ccgttgggca ggggtcctgg gccaccccg 840  
gcaggacgcg tggaccgagt gaccgtgggt tctgtgtggt gtcacctgcc agaccgcgg 900  
aagaagccac ctctttggag ggtgcgctct ctggcagcgg ccactcccac ccatccgtgg 960  
gccgccagca ccacgcgggc ccccatcca catcgcgggc accacgtccc tgggacacgc 1020  
5 cttgtccccc ggtgtacgcc gagaccaagc acttctctta ctctcaggc gacaaggagc 1080  
agctgcggcc ctcccttctta ctacgtcttc tgaggcccg cctgactggc gctcggaggc 1140  
tcgtggagac catctttctg ggttccaggc cctggatgcc agggactccc cgcaggttgc 1200  
ccgcctgcc ccagcgctac tggcaaatgc ggcctctgtt tctggagctg cttgggaacc 1260  
acgcgcagtg cccctacggg gtgtctctca agacgcactg cccgtctcga gctgcggtca 1320  
10 cccacgacgc cgggtgtctgt gcccgggaga agccccaggg ctctgtggcg gcccgcaggg 1380  
aggaggacac agaccctcgt cgctgtgtgc agctgtctcg ccagcacagc agccctcggc 1440  
agggtgtacg cttcgtgctg gcctgcctgc gccggctggt gcccacaggc ctctggggct 1500  
ccaggcacaa cgaacgcggc tctctcagga acaccaagaa gttcatctcc ctggggaagc 1560  
atgccaaagt ctgcctgcag gagctgacgt ggaagatgag cgtgcgggac tgcgctggc 1620  
15 tgcgcaggag cccaggggtt ggctgtgttc cggccgcaga gcacctctg cgtgaggaga 1680  
tcctggccaa gttcctgcac tggctgatga gtgtgtacgt cgtcagctg ctccaggctct 1740  
ctttttatgt cagcgagacc acgtttcaaa agaacaggct ctttttctac cgaagagtg 1800  
tctggagcaa gttgcaaaag attggaatca gacagcactt gaagagggtg cagctcggg 1860  
agctgtcggg agcagaggtc aggcagcatc ggaagccag gccgcctctg ctgacttcca 1920  
20 gactccgctt catccccaag cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980  
tgggagccag aacgttccgc agagaaaaa gggccgagcg tctcacctcg aggtggaag 2040  
cactgttccg cgtgtctaac tacgagcggg cgcggcgccc cggcctcctg ggcgctctg 2100  
tgctgggcct ggacgatata cacagggcct ggccgacctt cgtgtctcgt gtgcggggcc 2160  
aggacccgcc gcctgagctg tactttgtca aggtggatgt gacgggcccg tacgacacca 2220  
25 tccccaggga caggctcacg gaggtcatcg ccagcatcat caaacccag aacacgtact 2280  
gcgtgcgtcg gtatgccgtg gtccagaagg ccgccatgg gcacgtccgc aaggccttca 2340  
agagccacgt ctctaccttg acagacctcc agccgtacat gcgacagtgc gtggctcacc 2400  
tgacggagac cagcccgctg agggatgcgg tctcatcga gcagagctcc tccctgaatg 2460  
aggccagcag tggcctcttc gacgtcttcc tacgttctat gtgccaccac gccgtgcgca 2520  
30 tcaggggcaa gtccctacgtc cagtccagg ggatcccgca gggctccatc ctctccacgc 2580  
tgctctgcag cctgtgtctc ggcgacatgg agaacaagct gtttgcgggg attcggcggg 2640  
acgggctgct cctgcgtttg gtggatgatt tcttgttggg gacacctcac ctccaccagc 2700  
cgaaaacctt cctcaggacc ctggtccgag gtgtccctga gtatggctgc gtggtgaact 2760  
tgccgaagac agtggtgaa cttccctgtag aagacgaggg cctgggtggc acggctttt 2820  
35 ttcagatgcc gcccacggc ctattccctt ggtgcggcct gctgctggat acccgaccc 2880  
tggagggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcacc 2940  
tcaaccgcgg ctccaaggct gggaggaaaca tgcgtcgcaa actctttggg gtcttgcggc 3000  
tgaagtgtca cagcctgttt ctggatttgc aggtgaacag cctccagacg gtgtgcacca 3060  
acatctacaa gatcctcctg ctgcaggcgt acagggttca cgcattgtgt ctgcagctcc 3120  
40 catttcatca gcaagtgttg aagaacccca catttttctt gcgctgcatc tctgacacgg 3180  
cctccctctg ctactccatc ctgaaagcca agaacgcagg gatgtcgctg ggggccaagg 3240  
gcgcgcggcg cctctgcccc tccgagggcg tgcagtggct gtgccaccaa gcattcctgc 3300  
tcaagctgac tcgacaccgt gtcacctacg tgccactcct ggggtcactc aggcagagcc 3360  
agacgcagct gagtcggaag ctccccggga cgacgtgac tgccctggag gccgcagcca 3420  
45 accggcact gccctcagac ttcaagacca tccgtgactg atggccaccc gccacagcc 3480  
aggccgagag cagacaccag cagccctgtc acgcggggct ctacgtccca gggaggagg 3540  
ggcgcccccac acccaggccc gcaccgctgg gactctgagg cctgagttag tgtttggccg 3600  
aggcctgcat gtccggctga aggctgagtg tccggctgag gcctgagcga gtgtccagcc 3660  
aagggctgag tgtccagcac acctgccgtc ttcacttccc cacaggctgg cgctcggctc 3720  
50 caccacaggg ccagcttttc ctccacagga gcccggttcc cactccccac atagggaatag 3780  
tccatcccca gattcggcat tgttaccctc tcgcccctgc ctcttttgc ttccaccccc 3840  
accatccagg tggagacctc gagaaggacc ctgggagctc tgggaatttg gactgaccaa 3900  
agggtgtccc tgtacacagg cgaggacctc gcacctggat gggggctccc gtgggtcaaa 3960  
55 ttgggggggag gtgctgtggg agtaaaatag tgaatatatg agtttttcag ttttgaaaaa 4020  
aaaaaaaaa aaaaaaaaaa aa 4042

<210> 3  
<211> 11276  
<212> DNA  
60 <213> Homo sapiens

<400> 3  
acttgagccc aagagttcaa ggctacgggt agccatgatt gcaacaccac acgccagcct 60  
tggtgacaga atgagacctt gtctcaaaaa aaaaaaaaaa aattgaaata atataaagca 120  
65 tcttctcttg ccacagtgga acaaaaccag aaatcaacaa caagagggaat tttgaaact 180

atacaaacac atgaaaatta aacaatatatc ttctgaatga ccagtgaagtc aatgaagaaa 240  
 ttaaaaagga aattgaaaaa ttattttaag caaatgataa cggaacata acctctcaaa 300  
 acccacggtat tacagcaaaa gcagtgtctaa gaagggaagtt tatagctata agcagctaca 360  
 tcaaaaaagt agaaaaagcca ggcgcagtggt ctcagtcctg taatcccagc actttgggag 420  
 5 gccaaaggcgg gcagatcgcc tgaggtcagg agttcgagac cagcctgacc aacacagaga 480  
 aaccttgtcg ctactaaaaa taaaaaatta gctgggcatg gtggcacatg cctgtaattcc 540  
 cagctactcg ggaggtctgag gcaggataac cgcttgaacc caggaggtgg aggttgcggt 600  
 gagccgggat tgcgccattg gactccagcc tgggtaacaa gagtgaacc ctgtctcaag 660  
 10 aaaaaaaaaa aagtagaaaa acttaaaaaa acaacctaat gatgcacctt aaagaactag 720  
 aaaagcaaga gcaaaactaaa cctaaaattg gtaaaagaaa agaaataata aagatcagag 780  
 cagaaataaaa tgaaactgaa agataacaat acaaaagatc acaaaaatta aaagtgtggt 840  
 ttttgaagag ataaacaaaa ttgacaaacc ttgtccaga ctaagaaaaa aggaagaaag 900  
 acctaaataa ataaagtccag agatgaaaaa agagacatta caactgatac cacagaatt 960  
 caaaggatca cttagggcta ctatgagcaa ctgtactata ataaattgaa aaacctagaa 1020  
 15 aaaaatagata aattcctaga tgcatacaac ctaccaagat tgaacctaga agaaatccaa 1080  
 agcccaaca gaccaataac aataatggga ttaagccat aataaaaaagt ctctagcaa 1140  
 agagaagccc aggacccaat ggcttccctg ctggatttta ccaatcattt aaagaagaat 1200  
 gaattccaat cctactcaaa ctattctgaa aaatagagga aagaataactt ccaaaactca 1260  
 20 tctacatggc cagtattacc ctgattccaa aaccagacaa aaacacatca aaaacaaaca 1320  
 aacaaaaaaa cagaaagaaa gaaaactaca ggccaatatc cctgatgaat actgatacaa 1380  
 aaatcctcaa caaaacacta gcaaaacaaa ttaaaacaaa ccttcgaaag atcattcatt 1440  
 gtgatcaagt gggatttatt ccagggtatg aaggatggtt caacatagtc aaatcaatca 1500  
 atgtgatata tcatcccaac aaaatgaagt acaaaaacta tatgattatt tcactttatg 1560  
 cagaaaaagc atttgataaa attctgcacc ttctatgata aaaaacctca aaaaaccagg 1620  
 25 tatacaagaa acatacagcc caggcacagt ggctcacacc tgcgatccca gcactctggg 1680  
 aggcgaaggt gggatgattg cttgggcccga ggagtttgag actagcctgg gcaacaaaat 1740  
 gagacctggt ctacaaaaaa cttttttaa aaattagcca ggcatgatgg catatgcctg 1800  
 tagtcccagc tagtctggag gctgaggtgg gagaatcact taagcctagg aggtcgaggc 1860  
 tgcagtgaag catgaacatg tcactgtact ccagcctaga caacagaaca agacctcact 1920  
 30 gaataagaag aaggagaagg agaagggaaga agggaggagg aaggaggagg gaggagaagg 1980  
 agggagtgga ggagaagtg aaggggaagg ggaagggaaa gaggagaagg aagaacata 2040  
 tttcaacata ataaaagccc tatatgacag accgaggtag tattatgagg aaaaactgaa 2100  
 agcctttcct ctaagatctg gaaaatgaca agggccactt tccaccactg tgattcaaca 2160  
 tagtactaga agtccctagc agagcaatca gataagagaa agaaataaaa ggcatccaaa 2220  
 35 ctgggaagga agaatcaaa ttatcctgtt tgcagatgat atgatcttat atctggaata 2280  
 gacttaagac accactaaaa aactattaga gctgaaattt ggtacagcag gatacaaaat 2340  
 caatgtacaa aaatcagtag tatttctata ttccaacagc aaacaatctg aaaaagaac 2400  
 caaaaaagca gctacaaata aaattaaaca gctaggaatt aaccaaagaa gtgaaagatc 2460  
 40 tctacaatga aaactataaa atgttgataa aagaattga agagggcaca aaaaagaaa 2520  
 agatattcca tgttcataga ttggaagaat aaatactgtt aaaaatgtcca tactacccaa 2580  
 agcaatttac aaattcaatg caatccctat taaaatacta atgacgttct tcacagaaat 2640  
 agaagaaaca attctaatg ttgtacagaa ccacaaaaga cccagaatag ccaagctat 2700  
 cctgaccaa aagaacaaaa ctggaagcat cacattacct gacttcaaat tatactacaa 2760  
 45 agctatagta acccaaaacta catggtactg gcataaaaac agatgagaca tggaccagag 2820  
 gaacagaata gagaatccag aaaaaaatcc atgcatctac agtgaactca tttttgaca 2880  
 aggtgccaag aacatacttt ggggaaaaga taatctcttc aataaatggt gctggaggaa 2940  
 ctggatatcc atatgcaaaa taacaatact agaactctgt ctctcacat atacaaaagc 3000  
 aaatcaaaat ggtgaaagg cttaaatcta aaacctcaaa ctttgcaact actaaaagaa 3060  
 50 aacaccggag aaactctcca ggacattgga gtgggcaaag acttcttgag taattccctg 3120  
 caggcacagg caaccaagc aaaaacagac aaatgggac ataatcaagt aaaaagcttc 3180  
 tgcccagcaa aggaacaat caacaagag aagagacaac ccacagaatg ggagaatata 3240  
 tttgcaaaact attcatctaa caaggaatta ataaccagta tatataagga gctcaaaacta 3300  
 ctctataaga aaaaacaccta ataagctgat tttcaaaaat aagcaaaaga tctgggtaga 3360  
 55 catttctcaa aataagtcac acaaatggca aacaggcatc tgaaaatgtg ctcaacacca 3420  
 ctgatcatca gagaaatgca aatcaaaact actatgagag atcatctcat cccagtttaa 3480  
 atggctttta ttcaaaagac aggcataaac aaatgccagt gaggatgtgg ataaaaggaa 3540  
 acccttgagc actgttggtg ggaatggaaa ttgctaccac tatggagaac agtttgaaag 3600  
 ttcttcaaaa aactaaaaat aaagctacca tacagcaatc ccattgctag gtatatactc 3660  
 60 caaaaaagg aatcagtgta tcaacaagct atctccactc ccacatttac tgcagcactg 3720  
 ttcatagcag ccaaggtttg gaagcaacct cagtgtccat caacagacga atggaaaaag 3780  
 aaaatgtggt gcacatacac aatggagtac tacgcagcca taaaaaagaa tgagatcctg 3840  
 tcagttgcaa cagcatgggg ggcactggct agtatgttaa gtgaaataag ccaggcacag 3900  
 aaagacaaac tttctatgtt ctcccttact tgtgggagca aaaaataaaa caattgacat 3960  
 65 agaaaatagag gagaatgggt gttctagagg ggtgggggac aggggtgacta gagtcaacaa 4020  
 taatttattg tatgttttaa aataactaaa agagtataat tgggttgttt gtaacacaaa 4080

5 gaaaggataa atgcttgaag gtgacagata ccccatTTac cctgatgtga ttattacaca 4140  
 ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200  
 aattaaaaat ttaatggcca ggcacgggtg ctcatgtccg taatcccagc actttgggag 4260  
 gccgaggcgg gtggatcacc tgaggtcagg agtttgaac cagtctggcc accatgatga 4320  
 aacctgtct ctactaaaga tacaaaaatt agccaggcgt ggtaggcacat acctgtagtc 4380  
 ccaactactc aggaggctga gacaggagaa ttgcttgaac ctggggaggcg gaggttgacg 4440  
 tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500  
 acaaaaacaa aaaaaagaag attaaaattg taatttttat gtaccgtata aatatatact 4560  
 ctactatatt agaagttaa aattaaaaa attataaaag gtaattaaac acttaattcta 4620  
 10 aaataagaac aatgtatgtg gggtttctag cttctgaaga agtaaaagt attgcccacga 4680  
 tggcagaaat gtgaggaggg aacagtggaa gttactgttg ttagacgctc atactctctg 4740  
 taagtgaact aattttaacc aaagacaggc tgggagaagt taaaggaggc ttctataagc 4800  
 cctaaaacaa ctgctaataa tggtgaaagg taatctctat taattacca taattacaga 4860  
 tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtctc tcatttcagg 4920  
 15 tgcctttttt cttgtgtgct tggagatttt cgattgtgtg ttctgttttg gttaaactta 4980  
 atctgtatga atcctgaaac gaaaaatggg ggtgatttcc tccagaagaa ttagagtacc 5040  
 tggcaggaag cagggtggctc tgtggacctg agccacttca atcttcaagg gtctctggcc 5100  
 aagaccagg tgcaaggcag aggcctgatg acccgaggac aggaagctc ggatgggag 5160  
 20 gggcgatgag aagcctgcct cgttggtag cagcgcatga agtcccctta ttacgcttt 5220  
 gcaagattg ctctggatac catctggaaa aggcggccag cgggaatgca aggagtcaga 5280  
 agcctcctgc tcaaaaccag gccagcagct atggcgccca cccgggctg tgccagagg 5340  
 agaggagta aggcacctcg aagtatggct taaatctttt ttccacctga agcagtgaac 5400  
 aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaaacagaa agtcatgga 5460  
 gcaaccttct caaggga aaa cagacgccc gctctgcggg catttacctc ttctctctc 5520  
 25 cctctcttg cctcgcgggt ttctgatcgg gacagagtga ccccggtgga gcttctccga 5580  
 gcccgctg aggacctct tgc aaagggc tccacagacc ccccgctgg agagaggag 5640  
 ctgagcctg cttaataaca aactgggatg tggctggggg cggcacagca cggcggaat 5700  
 caaagactta attccatgag taatttcaac ctttcccat ccgaatggat ttggatttta 5760  
 30 tcttaataat ttcttaatt tcatcaaaata acattcagga ctgcagaaat ccaaggcgt 5820  
 aaaacaggaa ctgagctatg ttgtccagg tccaaggact taataaccat gttcagagg 5880  
 atttttcgcc ctgaagtact ttatttgggt ttcataagggt ggcttagggg gcaaggga 5940  
 gtacacgagg agaggcctg gcggcagggc tatgagcag gcagggccac cggggagaga 6000  
 35 tggcccgcc tgggaggct acagcaggac cactgaccgt cctccctgg agctgccaca 6060  
 ttgggcaacg cgaaggcggc cagcgtgctg gtgactcagg accccatacc ggcttctgg 6120  
 gcccccac actaaccag gaagtacag agctctgaac ccgtggaac gaacatgacc 6180  
 cttgcctgcc tgcttccctg ggtgggtcaa gggtaatgaa gtggtgtgca ggaatggcc 6240  
 atgtaaatta cacgactctg ctgatgggga ccgttctctc catcattatt catcttacc 6300  
 ccaaggact gaatgattcc agcaacttct tgggtgtga caagccatga caaaactcag 6360  
 40 tacaacacac actcttttac taggccacac gagcacggsc cacaccctg atataaag 6420  
 agtccaggag agatgaggct gctttcagcc accaggctgg ggtgacaaca gcggtgaac 6480  
 agtctgttcc tctagactag tagaccctgg caggcaactcc cccagattct agggcctggt 6540  
 tgtgtcttcc cgaggcgcc atctgcccgt gagactcagc ctggggtgcc acactgaggc 6600  
 cagccctgtc tccacacct ccgctccag gctcagctt ctccagcagc ttcttaaac 6660  
 45 ctgggtgggc cgtgttccag cgtactgtc tcacctgtcc cactgtgtct tgtctcagcg 6720  
 acgtagctc caggttctc cctcacatgg ggtgtctgtc tcttcccca acactcacat 6780  
 gcgttgaagg gagggattc tgcgctccc agactggctc ctctgagcct gaacctggct 6840  
 cgtggccccc gatgcagggt cctggcgtcc ggtgcacgc tgacctccat ttccaggcgc 6900  
 tccccgtctc ctgtcatctg ccggggcctg ccggtgtgtt cttctgttct tgtgtctct 6960  
 50 tccacgtcca gctgcgtgtg tctctgccg ctagggtctc ggggttttta taggcatagg 7020  
 acggggcgt ggtgggccag ggcgctctt ggaatgcaa catttgggtg tgaaagtagg 7080  
 agtgctgtc ctacactagg tccacgggca caggcctggg gatggagccc ccgcccagg 7140  
 ccgccccttc tctgccagc actttcctgc cccctcctt ctggaacaca gagtggcagt 7200  
 55 ttccacaagc actaagcatc ctcttcccaa aagaccagc attggcaccc ctggacattt 7260  
 gcccacagc cctgggaatt cacgtgacta cgcacatcat gtacacactc ccgtccacga 7320  
 ccgaccccg ctgttttatt ttaatagcta caaagcagg aaatccctgc taaaatgtcc 7380  
 tttacaacac tggttaaaca aacgggtcca tccgcacggt ggacagttcc tcacagtga 7440  
 gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgct gagtcaaac 7500  
 60 tgccacctcc atgggatagc tacgcaacat gctcaaaaag aaagaatttc acccatggc 7560  
 aggggagtg ttagggggtg taaggacggt gggggcggca gctgggggt actgacgca 7620  
 cctttacta aagccagttt cctggttctg atggtattg ctcagttatg ggagactaac 7680  
 cataggggag tggggatggg ggaacccgga ggctgtgcca tcttggccat gcccgagtg 7740  
 cctgggcagg ataatgtct agagatgccc acgtctgtat tccccaaac ctgtggacag 7800  
 65 aaccgcccgc gcccagggc ctttgcagg gtgactctcg tgaggacct gaggtctggg 7860  
 atcctcggg actacctgca gggccgaaaa gtaatccagg ggttctggga agaggcggc 7920  
 aggggggtca gaggggggca gcctcaggac gatggaggca gtcagtctga ggctgaaaag 7980

5 ggagggaggg cctcgagccc aggcctgcaa gcgcctccag aagctggaaa aagcggggaa 8040  
 gggaccctcc acggagcctg cagcaggaag gcacggctgg cccctagccc accagggccc 8100  
 atcgtggacc tccggcctcc gtgccatagg agggcactcg cgctgccctt ctgacatgaa 8160  
 gtgtgtgggg atttgcagaa gcaacaggaa acccatgcac tctgaatcta ggattatttc 8220  
 10 aaaaacaaag tttacagaaa catccaagga cagggtctga gtgcctccgg gcaaggcgag 8280  
 ggcaggcacg agtgatttta tttagctatt ttattttatt tacttacttt ctgagacaga 8340  
 gttatgtctt tgttgcccag gctggagtgc agcggcatga tcttggctca ctgcaacctc 8400  
 cgctcctctg gttcaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460  
 gtgcaccacc acacccggct aattttgtat ttttagtaga gatgggcttt caccatgttt 8520  
 15 gtcaagctga tctcaaaatc ctgacctcag gtgatccgcc caccctcagc tcccaaatg 8580  
 ctgggattac aggcacatgac cactgcacct ggccatttta accattttta aacttcccctg 8640  
 ggctcaagtc acacccactg gtaaggagtt catggagttc aatttcccct ttactcagga 8700  
 gttaccctcc tttgatattt tctgtaattc ttcgtagact ggggatacac cgtctcttga 8760  
 20 catattcaca gtttctgtga ccacctgtta tccccggga cccactgcag gggcagctgg 8820  
 15 gaggctgcag gcttcaggtc cacgtggggg tgccatctgc cagtagaaac ctgatgtaga 8880  
 atcaggggcg aagtgtggac actgtcctga atctcaatgt ctcagtgtgt gctgaaacat 8940  
 gtgaaaatta aagtccatcc ctctactct actgggattg agccccctcc ctatcccccc 9000  
 ccagggcgag agggagtctc ctactcctg tggagggaag aatgatactt tgttattttt 9060  
 25 cactgctggt actgaatcca ctgtttcatt tgttggtttg tttgttttgt tttgagaggc 9120  
 20 ggtttcactc ttgtgtctca ggctggaggg agtgcaatgg cgcgatcttg gcttactgca 9180  
 gcctctgctc cccaggttca agtgattctc ctgcttccgc ctcccatttg gctgggatta 9240  
 caggcacccg ccaccatgcc cagctaattt tttgtatttt tagtagagac gggggctggg 9300  
 ggggttcacc atgttgcca ggctgtgttc gaacttctga cctcagatga tccacctgcc 9360  
 30 tctgcctcct aaagtgtctg gattacaggt gtgagccacc atgcccagct cagaatttac 9420  
 25 tctgtttaga aacatctggg tctgaggtag gaagctcacc ccactcaagt gttgtgtgt 9480  
 tttaaagcaa tgatagaatt tttttattgt tgttagaaca ctcttgatgt ttacactgt 9540  
 gatgactaag acatcatcag cttttcaaag acacactaac tgcaccata atactgggg 9600  
 gtcttctggg tatcagcaat cttcattgaa tgcggggagg cgtttcctcg ccactgcacat 9660  
 35 ggtgttaatt actccagcat aatcttctgc ttccatttct tctcttccct cttttaaaat 9720  
 30 tgtgttttct atgttgctt ctctgcagag aaccagtgtg agctacaact taacttttgt 9780  
 tggaaacaaat tttccaaacc gccccttctg ccttagtgga gagacaattc acaaacacag 9840  
 ccccttaaaa aggccttaggg atcactaagg ggatttctag aagagcgacc tgtaatccta 9900  
 35 agtatttaca agacgaggct aacctccagc gagcgtgaca gcccagggag ggtgcgaggc 9960  
 ctgttcaaat gctagctcca taaataaagc aatttccctc ggcagtttct gaaagtagga 10020  
 aaggttacat ttaagggtgc gtttgttagc atttctagtgt ttgcccagct cagctacagc 10080  
 40 atccctgcga ggcctcgga gacccagaag tttctcgccc ccttagatcc aaacttgagc 10140  
 aaccggaggt ctggattcct gggaagtcct cagctgtcct gcggtgtgtc cggggcccca 10200  
 ggtctggagg ggaccagtgg ccgtgtggct tctactgctg ggctggaaat cgggctcctc 10260  
 45 agctctgcag tccgaggctt ggagccaggt gcctggacce cgaggctgcc ctccacctg 10320  
 40 tgcggggcgg atgtgaccag atgttgccct catctgccag acagagtgcc gggggccagg 10380  
 gtcaaggccg ttgtggctgg tgtgaggcgc ccggtgcgcg gccagcagga gcgcctggct 10440  
 ccatttccca ccccttctcg acgggaccgc cccggtgggt gattaaacaga tttggggctg 10500  
 50 tttgtctatg gtggggaccc ctgcgccctc gagaacctgc aaagagaaat gacgggctg 10560  
 45 tgtcaaggag cccaagtgcg ggggaagtgt tgcaggagg cactccggga ggtcccgcgt 10620  
 gcccgctccag ggagcaatgc gtccctcgggt tegtccccag ccgcgtctac gcgcctccgt 10680  
 cctccccctc acgtccggca ttctgtgtgc ccggagcccg acgccccgcg tccggacctg 10740  
 gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgccgca cgcacctgtt 10800  
 cccaggccct ccacatcatg gcccctccct cgggttacc cccagcctag gccgattcga 10860  
 55 cctctctccg ctggggccct cgctggcgtc cctgcacctt gggagcgca gcggcgcg 10920  
 50 ggcggggaaag cgcgccccc acccccgggt ccgcccggag cagctgcgct gtcggggcca 10980  
 ggccgggctc ccagtggatt cgcgggcaca gacgcccagg accgcgctcc ccacgtggcg 11040  
 gagggactgg ggaccgggg acccgtcctg ccccttcacc ttccagctcc gcctcctccg 11100  
 cgcggaaccc gcccgctcc gacccctccc gggctccccg cccagccccc tccgggccc 11160  
 60 cccagccctt ccccttctct tccgcggccc cgccctctcc tccggcgcg agtttcagge 11220  
 55 agcgtctgct cctgtctgc acgtgggaag ccctggcccc ggccaccccc gcgatg 11276

<210> 4  
 <211> 104  
 <212> DNA  
 60 <213> Homo sapiens

<400> 4  
 gtgggcctcc ccggggctcg cgtccggctg ggggttaggg cgccggggg gaaccagcga 60  
 65 catgaggaga gcagcgaggg cgaactcagg cgcttcccc gcag 104



<210> 5  
<211> 8616  
<212> DNA  
<213> Homo sapiens

5

<400> 5

gtgaggaggt ggtggccgtc gagggcccag gccccagagc tgaatgcagt aggggctcag 60  
aaaagggggc aggcagagcc ctggctcctcc tgtctccatc gtcacgtggg cacacgtggc 120  
ttttcgctca ggacgtcgag tggacacggg gatctctgcc tctgctctcc ctccgtgcca 180  
10 gtttcgataa acctacgagg ttacacctca cgttttgatg gacacgcggg ttccaggcgc 240  
cgaggccaga gcagtgaaca gagggaggctg ggcgcggcag tggagccggg ttgcccggca 300  
tggggagaag tgtctggaag cacagacgct ctggcgaggg tgcctgcagg ttacctataa 360  
tctcttctgc aatttcaagg gtgggaatga gaggtgggga cgagaacccc ctcttctctg 420  
gggtgggagg taagggtttt gcaggtgcac gtggtcagcc aatatgcagg tttgtgttta 480  
15 agatttaatt gtgtgttgac gggcagggtg ggtggctcac gccggtaatc ccagcacttt 540  
gggaagctga ggcaggtgga tcacctgagg tcaggagttt gagaccagcc tgaccaacat 600  
ggtgaaaccc tatctgtact aaaaatacaa aaattagctg ggcaggtgtg tgtgtgcctg 660  
taatcccagc tacttgggag gctgaggcag gagaatcact tgaaccagag aggcggaggc 720  
tgacgtgagc tgagattgtg ccattgtact ccagcctggg cgacaagagt gaaactctgt 780  
20 ctttaaaaaa aaaaagtgtt cgttgattgt gccaggacag ggtagaggga gggagataag 840  
actgttctcc agcacagatc ctgggtcccat ctttaggtat gaagagggcc acatgggagc 900  
agaggacagc agatggctcc acctgctgag gaagggacag tgtttgtggg tgttcagggg 960  
atgggtgctg tgggcccctg cgtgtcccca cctgttttt ctggatttga tgttgaggaa 1020  
cctccgctcc agcccccttt tgggtcccag tgtcccagg cctaccctg gcagctagaa 1080  
25 gaagtcccga ttccaccccc tccccacaaa ctcccaagac atgtaagact tccggccatg 1140  
cagacaagga gggtagacct cttggggctc tttttttct ttttttctt ttatgggtggc 1200  
aaaagtcata taacatgaga ttggcactcc taacaccgtt ttctgtgtac agtgcagaa 1260  
tgctaactcg gcggtgttta cagcagggtg cttgaaatgc tgcgtcttgc gtgactggaa 1320  
gtccctaccc atcgaacggc agctgcctca cactctgtgc ggctcagggt gaccacggc 1380  
30 agtcagataa gcgtcatgca acccagtttt gctttttgtg ctccagcttc cttcgttag 1440  
gagagtttga gttctctgat caggactctg cctgtcattg ctgttctctg acttcagatg 1500  
aggtcacaa ctgccccctg cttatgcagg gagttagggc tgggtccccg gtgtccctgt 1560  
cacgtgcagg gtgagtgagg cgttgccccc aggtgtccct gtcacgtgta gggtagtgta 1620  
ggcgcggccc cgggtgttcc ctgtcccgtg cagcgtgatt gaggtgtggc ccccggtgt 1680  
35 cctgtcacg tgtagggtga gtgaggcgcc atccccgggt gtccctgtca cgtgtagggt 1740  
gagttagggc tgggtcccgg gtgtccctgt cccgtgcagg gtgagtgagg cactgtcccc 1800  
gggtgtccct gtacgtgca ggtgagtgga ggcgcggctc cgggtgttcc ctctcagggt 1860  
tagggtgagt gaggcggcgc cccagggtgt cctgttcacg tgtagggtga gtgaggcacc 1920  
gtccctgggt gtccctccca ggtatagggt gagttaggca ctgtccccg gtgtccctgt 1980  
40 cacgtgcagg gtgagtgagg cgcggccccc ggggtgtccct ctccaggtgca gggtagtgta 2040  
ggcgtgttcc ctgggtgtcc ctgtctctgt tagggtgagt gaggtctgt cccaggtgt 2100  
ccttggcgtt tgcctcactg agcttctctc tgaatgtttg ctcttcttat agccacagct 2160  
gcgcgggtt cccattgcct ggttagatgg tgcaggcgca gtgctgggtc ccaagcctat 2220  
cttttctgat gctcggtctt tcttgggtac ctctccgttc cattttgtta cggggacacg 2280  
45 ggactgcagg ctctcgcttc ccgctgcca ggcactgcag ccacagcttc aggtccgctt 2340  
gcctctgttg ggcctggctt gctcaccacg tgcggccac atgcatgctg ccaatactcc 2400  
tctcccagct tgtctcatgc cagggtgga cctggggctg cctgtgtctg ctgccacgtg 2460  
ttgttggaaga catcccagaa aggggtctct gtgccctgaa ggaaagcaag tccccacagc 2520  
ccctcactt gtctgtttt ctcccaagct gccctctgtc ttggccccc ttgggtgggtg 2580  
50 gcaacgcttg tcaccttatt ctgggcacct gccgtcatt gcttaggctg ggctctgcct 2640  
ccagtcgccc cctcacatgg attgacgtcc agccacaggt tggagtgtct ctgtctgtct 2700  
cctgtctga gacccacgtg gagggccggg gtctccgcca gccttctgta gacttccctc 2760  
ttgggtctta gttttgaatt tcaactgatt acctctgacg tttctatctc tccattgtat 2820  
55 gctttttctt ggtttattct ttcatctctt ttctagcttc ttagttagt catgccttct 2880  
cctctaagtg ctgccttacc tgcacctgt gttttgatgt gaagtaatct caacatcagc 2940  
cactttcaag tgttcttaaa atacttcaaa gtgttaatac ttcttttaag tattcttatt 3000  
ctgtgatttt tttctttgtg cagcgtgtgt ttgacgtga aatcattttg atatcagtga 3060  
cttttaagta ttcttttagt tattctgtga ttcttttag cagttagtta ttgaaacct 3120  
gtttatgttc aagatatgta gagtatcaag atacgtagag tattttaagt tatcatttta 3180  
60 ttattgattt ctaactcagt tgtgtagtgg tctgtataat accaattatt tgaagttctg 3240  
ggagccttgc tttgtgatct agtgtgtgca tgggtttccag aactgtccat tgtaaatttg 3300  
acatcctgtc aatagtgggc atgcatgttc actatatcca gcttatttaag gtccagtgca 3360  
aagcttctgt ctccctctag atgcatgaaa ttccaagaag gaggcataag tccctcact 3420  
65 gggggatggg tctgttctat tcttctctgt tggtagcatt tatgtgaggc attgttaggt 3480  
gcatgcacgc ggtagaattt ttatcttctt gatgagtgaa tcttttgag acttctatgt 3540

ctctagtaaat cttagtaattc tttttttaa ttgctcttag tactgccaca ctgggcttct 3600  
 tttagattagt attttcctgc tgtgtctgtt ttctgccttt aatttatata tatatatata 3660  
 tttttttttt tttttgagaca gagtcttggg ctgtcgccca gggtagtgag agtgggtgtga 3720  
 tcacagggtca gtgtaacttt taccttctgg cctgagccgt cctctcacct cagcctcctg 3780  
 5 agtagctgga actgcagaca cgcaccgcta cacctggcta atttttaaat tttttctgga 3840  
 gacaggggtct tgctgtgttg cccagggtgg tctcaaaact ttggactcaa gggatccatc 3900  
 tacctcgggt tcccaaaagt ctgaattaca ggcagtagcc accatgtctg gcctaatttt 3960  
 caacactttt atattcttat agtgtgggta tgtcctgtta acagcatgta ggtgaatttc 4020  
 caatccagtc tgacagtcgt tgtttaactg gataacctga tttattttca tttttttgtc 4080  
 10 actagagacc cgcctgggtg actctgattc tccacttgcc tggtagcatg cctcgttccc 4140  
 ttgtttctca ccacctcttg ggttgccatg tgcgtttcct gccgagtggt tgttgatcct 4200  
 ctggttgccct cctggtcact gggcatttgc ttttatttct ctttgccttag tgttaccctc 4260  
 tgatcttttt attgtcgttg tttgtctttg tttatttgaga cagtctcact ctgtcaccca 4320  
 ggctggagtg taatggcaca atctcgggct actgcaacct ctgcctcctc ggttcaagca 4380  
 15 gtcttcattc ctcaacctca tgagtagctg ggattacagg cgcccaccac cagcctcctg 4440  
 taatttttgt atttttagta gagataggct ttcacctgtg tggccagggt ggtctcaaac 4500  
 tcctgacctc aagtgatctg cccgccttgg cctcccacag tgcctgggatt acaggtgcaa 4560  
 gccaccgtgc ccggcatacc ttgatctttt aaaaatgaagt ctgaaacatt gctacccttg 4620  
 20 acctgagcaa taagacctct agtgtatttt agctctggcc acccccagc ctgtgtgctg 4680  
 ttttccctgc tgacttagtt ctatctcagg catcttgaca ccccacaaag ctaagcatta 4740  
 ttaattattgt tttccgtgtt gagtgtttct gtagctttgc ccccgccctg cttttctctc 4800  
 ttgtttcccc gtctgtcttc tgtctcaggc ccgccgtctg gggctcccct ccttgcctct 4860  
 tgcgtgggtc ttctgtcttg ttattgtctg taaaccccag ctttacctgt gctggcctcc 4920  
 atggcatcta gcgagctccg gggagctctg cttatgatgc acagatgaag atgtggagac 4980  
 25 tcacaggag ggccgtcatc ttggcccgtg agtgtctgga gcaccacgtg gccagcgttc 5040  
 cttagccagt gagtgacagc aacgtccgtc cggcctgggt tcagcctgga aaaccccagg 5100  
 catgtcgggg tctgtgtggt ccgcggtgtc gagtttgaac tcgcgcaaac ctgcggtgtg 5160  
 gcgcagctc tgacggtgct gcctggcggg ggagtgtctg cttcctccct tctgcttggg 5220  
 aaccaggaca aaggatgagg ctccgagccg ttgtcgccca acaggagcat gacgtgagcc 5280  
 30 atgtggataa ttttaaaatt tctaggctgg gcgcggtggc tcacgcctgt aatcccagca 5340  
 ctttgggagg ccaaggcggtg tggatcacga ggtcaggagg tcgagaccat cctggccaac 5400  
 atgatgaaac cccatctgta ctaaaaacac aaaaattagc tggcgctggt ggccggtgct 5460  
 tgtaatccca gctactcggg aggctgagggc aggagaattg cttgaacctg ggagtggaa 5520  
 35 gttgcagtga gccgacattg caccactgca ctccagcctg gcaacacagc gagactctgt 5580  
 ctcaaaaaaa aaaaaaaaaa aaaaaaaaaa aattctagta gccacattaa aaaaatgaaa 5640  
 aagaaaaagg gaaattaatg taataataga ttttactgaa gccagcatg tccacacctc 5700  
 atcatcttag ggtgttattg gtgggagcat cactcacagg acatttgaca ttttttagc 5760  
 tttgtctgcy ggtatccctg ttaggttccc gtgcgtggcc atctcggcct ggacctgctg 5820  
 40 ggcttcccat ggccatggct gtgtaccag atggtgcagg tccgggatga ggtcgccagg 5880  
 ccctcagtga gctggatgtg cagtgtccgg atggtgcacg tctgggatga ggtcgccagg 5940  
 cctctgtgtg agctggatgt gtggtgtctg gatggtgacg gtcagggggt aggtctccag 6000  
 gccctcgggt agctggaggt atggagtcag gatgatgacg gtcgggggtg aggtcgccag 6060  
 45 ggcctcgtgt gagctggatg tgtggtgtct ggatggtgca ggtcagggggt gaggtctcca 6120  
 ggccctcggt aagctggagg tatggagtcc ggatgatgca ggtccgggggt gaggtcgcca 6180  
 gggccctgct tgagctggat gtgtggtgtc tggatgggtc aggtctgggg tgaggtcacc 6240  
 aggcctgctg gtgagctggg tgtgcggtgt ctggatgggt caggtctgga gtgaggtcgc 6300  
 cagacggtgc cagaccatgc ggtgagctgg atatgcggtg tccggatggt gcaggtctg 6360  
 50 ggtgaggttg ccaggccctg ctgtgagttg gatgtgggtg gtccggatgc tgcaggtcgc 6420  
 ggttgagggt accaggccct gctgtgagct ggatgtgtgg tgtctggatg gtgcaggtct 6480  
 ggggtgaggg tcgccaggcc cctgcttgtg agctggatgt tgggtgtctg gatggtgcag 6540  
 gtcttgagtg aggtcgccag gccctcgggt agctggatgt gcagtgtcca gatggtgcag 6600  
 gtcgggggtg aggtcgccag accctcgggt gagctggatg tgcggtgtct ggatggtgca 6660  
 55 ggtctggagt gaggtcgcca ggccctcggt gagctggatg tatggagtcc ggatggtgcc 6720  
 ggtccggggt gaggtcgcca gacctgctg tgagctggat gtgcggtgtc tggatggtac 6780  
 aggtctggag tgaggtcgcc agacctgct gtgagctgga tatgcggtgt ccggatggtg 6840  
 caggtcaggg gtgaggtctc caggccctcg gtgagctgga ggtatggagt ccggatgatg 6900  
 caggtccggg gtgaggtcgc caggccctgc tgtgaactgg atgtgcggcg tctggatggt 6960  
 gcaggtctgg ggtgtggtcg ccaggccctc ggtgagctgg aggtatggag tccggatgat 7020  
 60 gcaggtctcg ggtgaggtcg ccaggccctc cgtgagctg gatgtgcggc gtctggatcg 7080  
 tgcaggtctg ggtgtggtgc gccaggccct cgtgagctg gaggtatgga gtccggatga 7140  
 tgcaggtccg ggtgtgaggt gccaggccct gctgtgagct ggatgtgctg tatccggatg 7200  
 gtgcagctcg ggtgtgaggt gccaggccct gctgtgagct ggatgtgctg tatccggatg 7260  
 gtgcaggtct ggtgtgaggt caccaggccc tgcggtgagc tgggtgtgct gtgtccggtt 7320  
 65 gctgcaggtc cgggtgtgag tgcagggccc ctcggtgagc tggatgtgct gtgtcccgct 7380  
 gtccggatgg tgcaggtcca ggtgtgaggt gctaggccct tgggtggctg gatgtgcctg 7440

gtccggatgg tgcaggctctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgcgggt 7500  
 gctctcatgg tgcaggctctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgcgggt 7560  
 gtccggatgg tgcaggctccg gctgtgaggtc gccaggccctt gctgtgagct ggatgtgcgg 7620  
 tctctggatg gtgcaggctcc ggggtgaggt agccaaggcc ttcggtagagc tggatgtggg 7680  
 5 gtgtccggat ggtgcaggctc cggggtgagg tgcaggccctt ctgaggcttag ctggatatgc 7740  
 ggtgtccgga tgggtgcaggc cggggtgagg gtcaccaggc cctgaggctta gctggatgtg 7800  
 cgggtctctgg atgggtgcagg tccgggggtga ggtgcaggc cctgtctgtg agctggatgt 7860  
 gctgtatccg gatgggtgca gtcgggggtg aggtgcagg gccctgcagt gagctggatg 7920  
 tctgttatcc ggatgggtgca ggtctggcgt gaggtgcagg gccctgcagg ttagctggat 7980  
 10 atgagggtgtc ggatgggtgca ggtcgggggt gaggtcagca gccctgcagg ttagctggat 8040  
 gtgagggtgtc cggatgggtgc aggtctgggg tggaggtgcgc agggccctgc ggtgagctgg 8100  
 tctgtctgat ccggatgggtg cagggtccggg gtgaggtgcgc caggccctgc ggtgagctgg 8160  
 atgtgctgta tccggatgggt gcaggctctgg cgtgaggtgc ccaggccctg cggtagagctg 8220  
 gatgtgcagt gtacggatgg tgcaggctccg ggggtgaggtc gccaggccctt ggggtgggct 8280  
 15 gtatgtgtgt tgtctggatg gtgcaggctcc ggggtgaggt cggcaggccc tgggtgagc 8340  
 tggatgtgtg gtgtctggat gctgcaggctc cggggtgaggt tgcaggccctt cctggtagagc 8400  
 tggatgtgag gtgtcccggt gtcgaatgg tgcaggctcca gggtagaggtc gccaggccctt 8460  
 tggtagagctg gatgtgcccgt gtcaggatgg tgcaggctctg gggtagaggtc gccaggccctt 8520  
 20 tggtagagctg gatgtgcgggt gtcaggatgg tgcaggctccg gggtagaggtc accaggccctt 8580  
 cggtagatctg gatgtggcat gtccttctctg ttttaag 8616

<210> 6  
 <211> 2089  
 <212> DNA  
 25 <213> Homo sapiens

<400> 6  
 gtactgtatc cccacgccag gcctctgctt ctggaagtc tggaaaccca gcccgccctc 60  
 agcatgcgcc tgtctccact tgcctgtgct tccctggctg tgcagctctg ggtggggagc 120  
 30 cagggggccc gtacacaggc tgggtccagt ggattctgtg caaggcctct actgctctga 180  
 gctcacgttc tcttacttgt aaaatcagga gttgtgcca agtgggtctct aggggtctga 240  
 aagcagaagg gatataaatt agatggaaac actaccacta gcctcctctg ctttcccttg 300  
 gatgtgggtc tgattctctc tctctttttt tttctttttt tgagatggag tctcactctg 360  
 35 ttgcccaggc tggagtgcag tggcataatc ttggctcact gcaacctcca cctcctgggt 420  
 ttaagcgatt caccagcctc agcctcctaa gtgctgggga ttacaggcac ctgccaccac 480  
 gcctggctaa tttttgtact tttaggagag acgggggttc accatgttgg ccaggctgggt 540  
 ctcgaaactca tgacctcagg tgatccaccc accttggcct cccaaagtgc tgggtttaca 600  
 ggctaagcca ccgtgccag ccccgattc tcttttaatt catgctgttc tgatgaatc 660  
 40 ttaactctat tggatttagg tcatgagagg ataaatccc acccacttgg cgactcactg 720  
 cagggagcac ctgtgcaggg agcacctggg gataggagag ttccaccatg agctaacttc 780  
 taggtggctg catttgaaat gctgtgagat tttgtctgca atgttcggct gatgagagt 840  
 tgagattgtg acagattcaa gctggatttg catcagtgag ggacgggagc gctgggtctg 900  
 gagatgccag cctggctgag cccaggccat ggtattagct tctccgtgtc ccgcccaggc 960  
 45 tgactgtgga gggcttttagt cagaagatca gggcttcccc agctccctct cactctgag 1020  
 tccctggggg gccttgtgac acccctatgc ccaaatcagg atgtctgcag agggagctgg 1080  
 cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccgagc tgggtgctgg 1140  
 gccatttcc tgcattctgg ggagggtcag ggcttccct gtgggaacaa gttatacac 1200  
 aatgcacctt acttagactt tacacgtatt taatgggtgt cgacccaaca tggctatttg 1260  
 50 accagtattt tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtgggtgg 1320  
 ccccaagatg ctcttctgca ctactgggac tgtgttctg cctggggggc cttggaggcc 1380  
 cctcctcctt ggacagggtg ccgtgccttt tctactctgc tgggctctgc gctgcgggtc 1440  
 agggcaccag ctccggagca cccgaggccc cagtgtccac ggagtgccag gctgtcagcc 1500  
 acagatgccc aggtccagggt gtggcgcgtc cagccccctg gcccccatgg ggtgttttg 1560  
 55 gggaaaaggc caagggcaga ggtgtcagga gactgggtgg ctcatgagag ctgattctgc 1620  
 tcttggctg agctgcctgc agcagcctct cccgcccctt ccatctgaag ggtgtgggt 1680  
 ctttctacct gggggtcctg cctggggcca gccttgggt acccagtggt ctgtaccaga 1740  
 gggacaggca tctgtgtggt aggggcatgg gttcacgtgg cccagatgc agcctgggag 1800  
 caggctcccc ggtgctgag gtgggacagt caccctgggg gttgaccgcc ggactgggag 1860  
 60 tccccagggt tgactatagg accaggtgtc cagggtcccc gcaagtagag gggctctcag 1920  
 aggcgtctgg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgctgccc 1980  
 tgggtgcccgt gagccctcac tgagtcgggt ggggcttctg gcttcccctg agcttcccc 2040  
 tagtctgttg tctggctgag caagcctcct gaggggctct ctattgcag 2089

10

15

25

30

35

40

60

62

agctttattg agggagaccat atcttccctt gaactatggt cgggtttata gtaagtcagg 180  
 ggtgtggagg cctcccctgg gctccctgtt ctgtttcttc cactctgggg tcgtgtgggt 240  
 cctgtgtgtg tgtgtggcgg gtgggcagggt cttccaggcc tccttgtgtt cattggcctg 300  
 gatgtggccc tgggtacgct ccgtccctgg aattcccctg cgagtgtggg gctttctttc 360  
 5 tttctctttt tctttctttt tttttttttt tgataacaga gtctcgtctt tttttgcca 420  
 ggctggagtg gtttggcgtg atcttggctc actgcaacct gtgcttccct agttcaagca 480  
 attctcttgc ctccagcctc caagtagctg gaattatagg cgccaccac catgctgact 540  
 aatttttgta attttagtag agacgagggt tctccatgtt ggccaggctg gtctcgaact 600  
 10 cctgacctca ggtgatcctc ccacctcggc ctcccaaagt gctgggatga cagggtgtga 660  
 ccgcccgcgc ccggccgagac tcgcttccct cagcttccgt gagatctgca gcgatagctg 720  
 cctgcagcct tgggtgtgac aacctccgtt tctcttctcc aggtctcgtt aggggtcttt 780  
 ccatttcatg actctcttca cagaagagtt tcacgtgtgc tgatttccc gctgtttctt 840  
 gcgtaattgg tgtctgctgt ttatcgatgg cctccttcca tttctttag gctttgttta 900  
 15 ttgtgtttt tccggtcctt tgaaggaaaa gtctcgatta tggatgtttg aactttcttt 960  
 tctaacaag catctgaagt tgccgttttc cctctaaagc agggatcccc agggccctgg 1020  
 ctgtggagtg gcaccggtct ggggcctgtt aggaaccggg cgccagcggg gaggctagg 1080  
 ggggtgtggg gagccagcgt tcccgcctga gccccgcccc tctcagatca gcagtggcat 1140  
 gcggtgtctca gaggcgca caacctactg agaactgtgc gtgagagggg tctagattct 1200  
 20 tctgtcctta tgggaatcta atgctgatg atctgaggtg gaaccgtttg ctcccaaac 1260  
 catccccctt cccactgctg tctgtggaa aaatcgtctt ccacgaaacc agtccctggt 1320  
 accacaatgg ttggggaccc tgtgtctaaag acctgtctca gcagcctctc gtcagtgttg 1380  
 atatatggc ttctctgtgt tgagtccaga ataattacgg atttctgtga tgcttctccg 1440  
 cgacctcaga cccatgggct atttgtgggc gtgttgctgc ctctgggtt gggaaagggt 1500  
 25 caggcccat gtaccttctt gtactgcct tccaggttgg tttccagggt tgaatcgtac 1560  
 tctagtgtgt tttagccac ggccttgcg ccagctcctg ggggtctggg aacatgtctga 1620  
 agcacagagt caccgtgcgc gtcttttgat gctcacaag ctccaggcct cctgtgtccg 1680  
 tgttagtgtg tgtcacgtgc ctgtcacat cctgtcttgg ggacgcaggg gcttagcagg 1740  
 tcccgtagta aatgacaagc gtctggggg agtctgcaga ataggaggtg ggggtgcccg 1800  
 30 tctctctccc gcgtctcag actcttctcc tgcctgtgct gtggctgcac ctgcatccct 1860  
 gcaatccctc cagcactggg ctggagaggc ccgggagctc gagtgcact tgtgccactg 1920  
 gactgtggat ggcagtcggt caggggggtc tgatgtgtgg tgactgtgga tggcgggttg 1980  
 tcacaggggt ctgagtgtg gtgactgtgg atggcggctc tggggctctga tgtggtgact 2040  
 35 gtggatggcg gtcgtggggg ctgagtgtg gtgactgtgg atggcggctc tggggctctga tgtggtgact 2100  
 tgtggtgact gtggatggcg gtctgggggt ctgagtgtgg gactgtggat ggcggctcgt 2160  
 ggggtctgat tgggtgactgt ggtggcaggt cgtggggctt gatgtgtgtt gactgtggat 2220  
 ggcggtcgtg ggggtctgat tgggtgactgt ggtggcaggt cgtggggctt gatgtgtgtt 2280  
 gactgtggat ggcggctcgt ggggtctgat tgtggtgact gtggatggcg gtcgtgggg 2340  
 ctgagtgtg gtgactgtgg atggcggctc tggggctctga tgtgtgtga ctgtggatgg 2400  
 40 cgggtcgtgg gtctgatgtg gtgactgtgg atggcggctc tggggctctga tgtgtgtga ctgtggatgg 2460  
 ctgtggatgg tgatcggctc caggggtctg atgtgtgggt actgtggatg gcggtcgtgg 2520  
 ggtctgatgt gtggtgactg tggatggtga tccgtcacag ggggtctgat tgtggtgact 2580  
 gtggatggcg gtcgtgggggt ctgagtgtg gtgactgtgg atggcgggtt gtcgggggg 2640  
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtctgatgtg tgggtgactgt 2700  
 45 ggtggcgggt cgtggggctt gatgtgtgtt gactgtggat ggcggctcgt ggggtctgat 2760  
 tgtggtgact gtggatggcg gtcgtgggggt ctgagtgtgt gactgtggat ggcggctcgt 2820  
 ggggtctgat tgggtgactgt ggtggcgggt cgtggggctt gatgtgtgtt gactgtggat 2880  
 ggcgggttgt cccgggggtc tgatgtgtgg tgactgtgga tggcggctcgt ggggtctgat 2940  
 50 gtggtgactg tggatggcag tctgggggtc tgatgtgtgg tgactgtgga tggcggctcgt 3000  
 ggggtctgat gtgtgggtgac tgtggatggc ggtcgtgggg tctgatgtgt ggtgactgtg 3060  
 gatggcgggtc gtgggggtctg atgtgtgggt actgtggatg gcggtcgtgg ggtctgatgt 3120  
 ggtgactgtg gatggcgggtc gtgggggtctg atgtgtgggt actgtggatg gtgatcggtc 3180  
 acaggggtct gatgtgtgtt gactgtggat ggcggctcgt ggggtctgat tgtggtgact 3240  
 gtggatggcg gtcgtgggggt ctgagtgtgt gactgtggat ggcggctcgt ggggtctgat 3300  
 55 tgtggtgact gtggatggcg gtcgtgggggt ctgagtgtgt gtgactgtgg atggcagctc 3360  
 gtcacagggg tctgatgtgt ggtgactgtg gatggcgggtc gtgggggtctg atgtgtgtgt 3420  
 actgtggatg ggcgtcgtgg ggtctgatgt gtggtgactg tggatggcg tctgtgggtc 3480  
 tgatgtgtgg tgactgtgga tggcgggtcgt ggggtctgat tgggtgactg tggatgggtga 3540  
 tccgtcacag ggggtctgat tgtggtagct gcaggtggag tcccagggtg gtctgtagct 3600  
 60 accttgcgtc ctccggcccc cggccccctt tccccaaaca gaagcttccc aggcgtcttc 3660  
 tgggtctcat cccggcatcg ggcttggccg cagggtccaca cgtcctgac ggaagaaaca 3720  
 agtgcaccag tctggccggg gcaggccaca tttgtggctc atgcctctc ctctgccggc 3780  
 ag 3782

12 / 18

<210> 11  
<211> 980  
<212> DNA  
<213> Homo sapiens

5

<400> 11  
gtctcgggcac tgccttgcag ggttgggcac ggactcccag cagtgggtcc tccccgggc 60  
aatcactggg ctcacgaccg gacagactgt tggccctggg gggcagtggg gggaaatgagc 120  
tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180  
10 tgcgacagct gctgcattca ggcacctgct cacgtttgac tgcgcgccct ctctccagtt 240  
ccgcagtgcc ttgtttcatg atttgcataa tgtcttctct gccagttttg atctttgaggc 300  
caaaggaaaag gtgtccccct cctttaggag ggcaggccat gtctgagccg tgtcctgccc 360  
agctggcccc tcagtgtctg gtctgaggcc aaaggaaaag tgtccccctt cttaggaggga 420  
cgggcccgtgt ttgagccacg ccccgctgag cgggcccctc agtgctgggt ctgtccacgt 480  
15 gggccctgtg ccctttgcag atgtgtctct tccacgtggc cctgtggctc ttgacagatg 540  
cctgttagca cttgctcggc tctaggggac agtcgtgtcc accgcatgag gctcagagac 600  
ctctgggcga atttcccttg cttccagggt gggggtggag gtggcctggg ctgctgggac 660  
ccagaccctg tgcggcgag ctgggcagca actcctggat cacatatgcc atccggggcca 720  
cggctgggctg tgtgggtgtg agcccagctg gacccacagg tggccagag gagacgttct 780  
20 gtgtcacaca ctctgcctaa gcccattgtg gtctgcagag actcggcccg gccagccac 840  
gatggccctg cattccagcc cagccccgca cttcatcaca aacactgacc ccaaaaggga 900  
cggagggtct tggccacgtg gtcttcgctg tctcagcacc caccggctca ctcccatgtg 960  
tctcccgtct gcttttcgag 980

25

<210> 12  
<211> 2485  
<212> DNA  
<213> Homo sapiens

30

<400> 12  
gtgagtcagg tggccagggt ccattgccct gcgggtggct gggcgggctg gcagggtctc 60  
tgctcacctc tctcctgccc cttccccact gnccttctgc cggggggcac cagagtctcc 120  
tttcttgccc cccgccccct ccggtctctg ggtgacagg tcccagggcc ccggaaacat 180  
35 ggctcggctt gcggcagccg gagcggagca ggtgccacac gaggcctgga aatggcaagc 240  
gggtgtgtga gtgtctcctg cgtggaggac gaggggcggg ggggtgtgtc gggctcaggtg 300  
tgcgcccagc gtttgagcct cagccttctc agctccaagt tactactgac cctggacacc 360  
cggctctcac acgcttctat ctctctctcc cgatacaaaa ggattttatc cgattctcat 420  
tccctgctcc gtctgtgac ccccgcgagg gcgcgggctc tctctctctg tactagattt 480  
40 cccatctgga aagtgcgggg ttgaccgtgt agtttctctc tctcgggggg cctgtggtg 540  
ccatggggca ggcggcctgg gagagctgcc gtcacacagc cactgggtga gccacactca 600  
cgggtgtaga gccacagtc ctggtgccac atcacgtcct ctggatttta agtaaaacca 660  
cacacctccc ggcaggcacc tgcctgcgac cctgtgtgtg cctggggaga gtggtagcac 720  
ggaggaaaatt cgtgcacat caaggtcatc agcaaggta tccgcagta ggtggaacgt 780  
ggaggcctct ctctgggac gtctccagcg gataaaggac tgtgcacagc ttcggaagct 840  
45 tttattttaa aataatacta ttaattattg cattataagt aatcactaat ggtatcagca 900  
attataatat ttattaaagt ataattagaa atattaagta gtacacacgt tctggaaaaa 960  
cacaaattgc acatggcagc agagtgaatt ttggccgagg gacacgtgtg cacatgtgtg 1020  
taagcgcccc ccaggcccc agaattcgtc gacaaagtca cctccccaga gaagccacca 1080  
cgggcccctc tctgtgtctg gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140  
50 gtggcagggc tttggggaat gtgaggtgat gactgcgtcc tcatgccctg acagacagga 1200  
ggtgactgtg tctgtcctgt ccttaggaca cggacaggcc cgaagctcta gtccccatcg 1260  
tgtgtccagt tggcctctga ataaaaacgt cttcaaaacc tgttgcccga aaaactaaga 1320  
acagagagag tttcccatcc catgtgctca caggggcgta tctgcttgcg ttgactcgtc 1380  
gggctggccg gactcctaga gttggtgctg gtgcttctgt gcaaaaagtg cagtctctct 1440  
55 gcccactact gtgatattct caccagcaag gaaagcctct tttcttttct ttttttttt 1500  
ttttttgaga cggaacgtca ctgttctctg cctgggcttg agtgacgtg cgcatctca 1560  
actcactgca acctccgct cccgggttcc agcatttctc ctgctctcag cttcccgagca 1620  
gctgagatta caggcaccca cccctgcgc ctggctaatt tttgtatttt tagtagagag 1680  
gggtttttgc catgttggcc aggtgtgtct cgaactcctg acctcaggtg atccaccac 1740  
60 ctcggcctcc caaagtgtg ggattacagg tgtgagccat cagcgccagc cggaaaagcct 1800  
ctttttaagg tgaccaccta tagcgttctc cgaaaaatac aggtcttctt ttgacgtag 1860  
gtctgaagcg tctcttagca acaggagtg cgtcctgtgg gctctgggga tggctgaggg 1920  
ctcgtggcca gccatgcct ctgtgtgcac ctttaggttc cagggggcta tttctgtctc 1980  
actgtttgtc tgaaaaacga cccctggcat ccttgttttg agagtctctg cttctcgttg 2040  
65 gtcatgctga aactaggggc aaggttgtat ccgttggcgc gcagcggtca catgtagggt 2100

13 / 18

catgagtcctt tcaccgtgga caaatccctt gaaaaaaaaa aaaggagctcc ggtaaagcat 2160  
 tcattccggg tcaagtgtct ggttctgtga ataaactcta agatttaaga aaccttaatg 2220  
 aaagaaaacc ttgatgatc agagcaagga tgtggtcaca cctgtggctg gatctgttcc 2280  
 agccgcccga gtgcatgggt agagtgggga gcagggtattg tttgttcaga ggtctcatct 2340  
 5 ggtagtgttc tgagggtgtt gccggctgaa tggtagacgt gtcgtttgtg tgtatgaggt 2400  
 tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgcccgtc 2460  
 tctcacctgt gtcttcccgc cccag 2485

<210> 13  
 10 <211> 1984  
 <212> DNA  
 <213> Homo sapiens

<400> 13  
 15 gtgaggcttc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tgaatgcattc 60  
 agtgttaata ttcttgggtc tctggagacc atgactgtct tgccttgagg aaccagacaa 120  
 ggttgcagcc ccttcttgggt atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180  
 gggctccacg caggctctgt ccagcggcca tgtccagagg cctcagggct cagcaggcgg 240  
 20 gagggccgct gccctgcatt atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300  
 ctgtcacgtc acccaggttc cgttagggtc cttggggaga tggggctggt gcagcctgag 360  
 gccccacatc tcccagcagg ccttcgacag gtggcctgga ctgggcgcct cttcagccca 420  
 ttgcccattc cacttgcatt gggctctaac ccaaggacgc acacacctaa atatcgtgcc 480  
 aacctaatgt ggttcaactc agctggcttt tattgacagc agttactttt ttttttttaa 540  
 tacttttaagt tctagggtag atgtgcacga cgtgcagggt agttacatat gtatacatgt 600  
 25 gccatgttgg tgtgtgtcac ccattaactc atcatttaca ttaggtatat ctccaatatg 660  
 tatccctccc cactcccccc atcccatgac aggccctggg gtgtgatgtt cccacccttg 720  
 tgtccaagtg ttctcattgt tcagttccca cctgtgagtg agaacatgtg gtgtttggtt 780  
 ttctttctct gcaatagttt gctcagagtg atgggtttcca gcttcgtcca tgtccctaca 840  
 aaggacatga actcatcctt ttttatgact gcatagttat ccgtgggtga tatgtgccac 900  
 30 attttcttaa tccagtctat catcgatgga catttgggtt ggttgcaagt ctttctact 960  
 gtgaatagtg ccgcaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020  
 tcctttgggt atataccag taatgggatg gctgggtcaa atggtatttc tagttctaga 1080  
 tccttgagga atcaccacac tgccttccac aatgggtgaa ctagtttaca cttccagcaa 1140  
 35 cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt ttatgaaaa 1200  
 tagtatcact gaacaagcag acagttagtg aaggatgctg cagggaagcct gcaggccaca 1260  
 cagccatttc tctcgaagac tccgggtttt tctgtgcat cttttgaaac tctagctcca 1320  
 attatagcat gtacagtgga tcaaggttct tcttcattaa ggttcaagtt ctatattgaa 1380  
 ataagtattat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttggg 1440  
 40 ggaagtgtc ctgcagctgg cggcacactg gtcagccctc tgggacagga tacctctggc 1500  
 ccattggcat ggggcgctgg gcttgggctt gagggtcaca cagtgcacca tgcccagctt 1560  
 cctgtggata ggtctgggt ctcggatcat gctgaggacc acagctgcca tgcgtgtaaa 1620  
 gggcaccacg tggctcagag ggggcgaggt tcccagcccc agctttctta ccgtcttcag 1680  
 ttatttttcc ctaagagtct gagaagtggg gccgcgctg atggccttcg ttcgtcttca 1740  
 45 gctggcacag aattgcacaa gctgatggta aacactgagt acttataatg aatgaggaat 1800  
 tgctgtagca gtttaactgt gagagctcgt ctgttggaaa gaaatttaag tttttcattt 1860  
 aaccgctttg gagaatgtta ctttatttat ggctgtgtaa attgtttgac attcagctcc 1920  
 tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcccttatt 1980  
 ttag 1984

50 <210> 14  
 <211> 1871  
 <212> DNA  
 <213> Homo sapiens

55 <400> 14  
 gtgaggcccg tgccgtgtgt ctgtggggac ctccacagcc tgtgggcttt gcagttgagc 60  
 ccccggtgtc ctgccccctg caccgcagcg ttgtctctgc caagtccctc ctctctgccg 120  
 gtgctggatc cgcaagagca gaggcgcttg gccgtgcacc caggcctggg ggcgcagggg 180  
 60 caccttcggg agggagtggt tacctgtcag gccctggtcc tgcaagagac caccaggtt 240  
 acacacgtgg tgagtgcagg cggtagacct gctcctgctg ctctttggaa agtcaagagt 300  
 ggcggtctct ggggccccag tgagaccccc aggagctgtg cacaggccct gcagggccga 360  
 ggcgagcagc tcctccccag ggtgcacctg agcctgcgga gagcaggagc tgcctgagtga 420  
 gctggcccac agcgttcgct gcggtcacgt tcctgctggg ggttgtttgg gatcggtggg 480  
 65 agaatttggg tttgtctagt gctgtgtctt tgaaccacgg agatggctag gagggtgtt 540  
 cagagttgat ttttgtgaat caaactaaaa tcaggcacag gggacctggc ctgagcacag 600

14 / 18

5 gggattgtcc aatgtgggtcc ccctcaagggt cgtccacacag agccgggtggg cttgttttaa 660  
 agtgcgattt gacgagggac gagaaacctt gaaagctgta aagggaaacc tcagaaaatg 720  
 tggccgccag ggggtggtttc aggtgctttg ctgggctgtg tttgtgaaaa cccattttgga 780  
 cccgccctcc aagtcacccc tccagggtcca ccctccagggt ccgccctggg ctgggggtat 840  
 gcctggcggt ccttgtgccc cagcccggag cacagcaggc tgtgcacatt taaatccact 900  
 aagattcact cggggggagc ccagggtcca agcaactgag ggctcaggag tccctgaggct 960  
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagtccc tgaggggtgt 1020  
 gggccaggag gtggctcaga gtgtatgtt ggtcccacc gggggcagaa ctctgtctct 1080  
 gatgagtcgg cagccatgta acaggaagggt gtggccacag ggagctggga atgcaccagg 1140  
 10 ggagctgcgc agctggccga ggtcccagggt ccaggccaca ggaagggcag ggggacgccc 1200  
 gggggccacag cagaggccgc aggaagggaa ggggatgcc aggccagagc agaggctacc 1260  
 gggcacagggt gggctccctg agctgggtga gcgaggctca tgactcggcg agggaaacct 1320  
 cttgacgtga agctgacgac tgggtgtggt cagctcacag cccagccagg tcccgccct 1380  
 gaggcaggaac tcagaaacct cccctttgtc taaagcacag cagatgcctt cagggcatct 1440  
 15 aggagaaaaac aggcataagc gttgagaaac gtcttaaaag aagggtggat ggtggcaatt 1500  
 tcttgtccag atttttagtct gccccggacc acagatgagt ctataacgggt atttgtgtgt 1560  
 tgccatgggg acacatgaga tggaccatca cagaggccac tggggctgca cctcccatct 1620  
 gagtcctggc tgtcccgggt ccaggccagg ttcttgcatg ctcacctacc tgcctgtccc 1680  
 gggagacagg gaaagcaccc cgaagtctgg agcagggtgt ggtccagggt cctcagagct 1740  
 20 cctgccaggc ccagcacctt gctccaaatc accactcttc tggggttttc caaagcattt 1800  
 aacaagggtg tcagggttacc tctgggtgta cggccccgca tctcgggggt gacattgtccc 1860  
 ctctgctta g 1871

25 <210> 15  
 <211> 1801  
 <212> DNA  
 <213> Homo sapiens

30 <400> 15  
 gtgagcgac ctggccggaa gtggagcctg tgccccggtg gggcagggtg tctgagcagg 60  
 ccgttgctgc cactctgtct tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120  
 gggccacagggt tgccccctgt cccatctggg gctgagcaga aatgcattct tctgtgggag 180  
 tgaggggtgct cacaacggga gcagttttct gtgctatttt ggtataaagg aatgggtgac 240  
 cagacctggg tgcactgagg tgtcttcaga aagcagctct gatccgaacc caagacgccc 300  
 35 gggccctgct gggcgtagt ctctcaaac cgaacacagg gggcctgtgt ggcctgagtc 360  
 cctctgaacc cgagacctg gggccctgct gggcgtagt ctctccgaac ccagagactt 420  
 cagggccctt tggggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480  
 agtctacagg atgccatgag ttcatgatca cgtgtgaccc atcaggggac agggccatgg 540  
 tgtggggggg gtctctaca aattctgggg tctgtttcc ccagagcccc agagctcaag 600  
 40 gccccgtctc aggtctagac acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660  
 ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaagggt gctcacacct 720  
 ataatcccag cactttggga gggcgagggt ggtggatcac ttgaggccag gagtttgagg 780  
 ccaacctaac caacatagtg aatttccatt tctacttaaa aaatacaaaa attagcttgg 840  
 cctggtggca cagcctgta gtccccgcta tgcgggaggc tgaggcagga gaatcatttg 900  
 45 aaccaggag gcagagggtg cagtgcagcg agatcacacc actgcactcc agcctgggca 960  
 acagagttag acttcatctt aaaaaaaaaa aaaaaagtat cagcattcca aaaccatagt 1020  
 ggacagggtg ttttttattc tgtccttcga taatatctac tgggtgctgtg ctgagggccg 1080  
 gaactggggg tgccttcctc tgaaaggcac accttcatgg gaagagaaat aagtgtgtaa 1140  
 tgggtgttaa accagagggt taaactgggg tccgtgctgt ctgagttaac agtccagatc 1200  
 50 tggactttgc ctctttccag aatgctccct ggggtttgct tcatggggga gcagcagggt 1260  
 tggacacctt cgtgagggg gagcagcagg tgcagacgcc ctcatgatgg gggagtgcca 1320  
 ggtgcagaca ccctgtgcca tgggtcccag catgtccctg ttgcagctcc ctccccaca 1380  
 ggatgccgggt ctctgtgctt cccacagtc cctgcttccc tctcacagcc ttacctgggt 1440  
 ctggcctcca ctggctttgt ctgcatgatt tccacatttc ctgggctccc agcacctctt 1500  
 55 cgcctctccc aggcacctct gcagtgtggt ccataccagt cagctgtgaa ctgtccactg 1560  
 cttattttgc tcccatagaa atgtattttt taggacaggc acccctgggt ccagcctctg 1620  
 gcacagcatc agtgaatgtt attgaaggac aaaggacaga caaacaatc aggaaaatg 1680  
 gtctctctta aacacattgc aaagccacag aggttagtgc aggatgggtg ggcacagggt 1740  
 catcagatgt ggggtccaat ccagaatatt ctgtgctccc aaaggccact tggtcagagt 1800  
 60 gtgtgcttgc agaggtggct ctaaaagctc agcagtgagg gcagtggttc gccatctca 1860  
 ggggtgaactc acatcctctg tgtctgaagt atacagcaga ggcttgaaag gcattctggga 1920  
 gaagaaaaca ggcataatga ttaagaaaag tgaataaagg aaagtgttaa gatgggaatt 1980  
 tctttgtcca gatttttagt tcaccaacca cagctcagat ggtagaatgt ggtcagaact 2040  
 gatggacaga acaatagaac aaaacggaag ccctatctct cagaacgtg tgttaatgtg 2100  
 65 gtatgtggca cagctgatgg aaaagagagt gtgtgtgtaa tttttttttc tgagaaaact 2160



15 / 18

5 gactggaagc aaataagttg tgtctttaca gcatatacca gagcagattc taggtagaag 2220  
aggagacaca tgcaaacac accagcaaca gaaataaaac aaaagactca aaggggaagg 2280  
aggtgaacgt tccctgggtt ggtgttgggg aaggacacac agggaggcgg atgaaaccag 2340  
tgaggcaacg ggcattgctt tcaatgcaga gaaactcagc ttgcctgagc cacagtgaag 2400  
atggccattc cctggagcgt ttgtgcacgt gatttattta aggcgccctg tgaggctctg 2460  
cacattcatc ctctcacttt gtctctctaa ccacctgaga ggtagaggag gaaaggctcc 2520  
aggggagcag ccgcccttgg tcacccagct ggcaaaagggc atgcattgatt gcagcctggc 2580  
ctctctctcc gggggccctt ctctgcccga ggacccca caagtcagac ccataggctc 2640  
aggggtgagc ggagcccaag gtcgtgttgg ggatggctgt gaaagaagaa atggacgtct 2700  
10 gatgcacact tgggaaggct ctaccagcag cgtcaaaaga atgcattgta aactgacagc 2760  
gagaccatc cctcaaaaga acgcacgtga aactgatggc gagacctgtc cccatccctc 2820  
atgctggctc cttttctggg cttgccaaaga gccagcatca ggttgaggca agctggaaag 2880  
acttttctgg aaagcagctt gtttgcattg aagtcctcac aatgtcctgt gtcttccag 2940  
taattccact tctgaagtga ccagacatta tcacgggtct tatttaccat ttccagtggt 3000  
15 ccaggcaggg ggacttgcca cagcaagtc cgaacctgcc caaatacagg gctaaggaga 3060  
tattatgcat cacaacactt gctctgccat taaacatttt tcaagaattt ttgaagaat 3120  
gtttaatggc acaaacgtt tatttcaatg tagcagtggt caaagctgga tgtaaaagaa 3180  
cacacccag gagcctgccc tgaatgtcat gtgtgttcat ctttgacat ggacatcat 3240  
gggcagtgag tgggtgtgag gccctggagg acatcggtgg gatgcctcca tctgccccct 3300  
20 ctggagacac catgtgtgcc acgtgcactc actggagccc tgtttagctg gtgccacctg 3360  
gctcttccat cctcgagatt caaacacagt gagattcccc acgcccactc cagtgtctct 3420  
ccacaaaaaa cctgagtcac acctgtgttc actcgaggga cgcccgagg ccagggtccc 3480  
acagcttatt atgtgttttt ggtcgagtta tgtgcagatc tcatcagggc agatgatgag 3540  
25 tgcacaaaaa cggccgtgag aggtttggat acactcaaca tcaatagcca ggtcctggtg 3600  
gagtttggtc atgcagagtc tggatggcat gtagcatttg gagtccatgg agtgagcacc 3660  
cagccccctc gggctgcagc gcatgcccga ggagggacaa ggaagcgga ggaagcgagg 3720  
aggtcttttg gagcaagctt tgcaggaggg ggctgggtgt ggggcaggca cctgtgtctg 3780  
acattccccct ctgtgtctca g 3801

30 <210> 16  
<211> 880  
<212> DNA  
<213> Homo sapiens

35 <400> 16  
gtgagcaggc tgatgggtcag cacagagttc agagtccagg aggtgtgtgc gcaagtatgt 60  
gtgtgtgtgt gtgcgcgcgt gcctgcaagg ctgatgggtg ctggctgcac gtaagagtgc 120  
acatgtacgc atatacacgt gagcacatca atgtgtgcat gtgtgtacat gaaggcatgg 180  
40 cagtgtgtgc acaggtgtgc aaggggacaa gtgtgtgcac atgcgaatgc acacctgaca 240  
tgcatgtgtg ttcgtgcaca gtcgtgtggg cattcacgtg aggtgcatgc gtgtgggtgt 300  
gcagtgtgag tagcatgtgt gcacataaca tgtattgagg ggtcctcgtg ttccccccg 360  
taggtcctca gcaccagtg cactccttac aggatgagac ggggtcccg gccttggtgg 420  
gctgaggctc tgaagctgca gccctgaggg cattgtccca tctgggcac cgcgtccact 480  
50 cctctccttg tgggcttctg tgtccactcc cctctccttg tgggcattta catccactcc 540  
actcctcttc tctgtgggc atccgcgtcc actccccctc tctgtgggca tctgcgtcca 600  
cctccccctc ctgtgggcat ttgcgtccac tccctctcct ggttcccttc tgtcttggtc 660  
gagcctcggg ggcaggcaga tgacacagag tcttgactcg cccagggtgg ttcgcagctg 720  
ccgggtgagg gccaggccgg atttcaactg gaagagggat agtttcttgt caaaatgttc 780  
ctctttcttg ttccatctga atggatgata aagcaaaaag taaaaactta aaatcccaga 840  
50 gaggtttcta cgttttctca ctcttctctg gcgactctag 880

<210> 17  
<211> 3186  
<212> DNA  
55 <213> Homo sapiens

<400> 17  
gtgagccgcc accaaggggt gcaggcccag cctccaggga ccctccgcgc tctgctcacc 60  
tctgaccccg ggtttcacct tggaaactcct gggttttagg ggcaaggaaat gtcttacgtt 120  
60 ttcagtgggt ctgctgcctg tgcacagttc tgttcgcgtg gctctgtgca aagcacctgt 180  
tctccatctc tgggtagtgg taggagccgg tgtggcccca ggtgtcccca ctgtgctgt 240  
gcactggccg tgggacgtca tggaggccat cccaggggcag caggggcatg gggtaaaag 300  
atgtttatgg ggagtcctag cagaggaggc tgggaagggt tctgaacagt agatgggaga 360  
tcagatgccc ggaggtattg ggggtctcagc aaagaggggc gaggtgggtg caggtgagg 420  
65 tgcgtggccc caccgccggg aagggtgcagc agagctgtgg ctccccacac agcccgcca 480

16 / 18

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60

```

gcacctgtgc tctgggcatg gctgtgtctc tggaaacttc cctgtcctgg ctgggtcagg 540
gggtccctcg ccaagaatcg acaactttat cacagaggga agggccaatc tgtggaggcc 600
acagggccag cttctgctcg gagttagggc aggtggtggc acaagcctcg gggctgtacc 660
aaagggcagt cgggcaccac agggccgggg ctcacactca acaggcctcc cgagccactg 720
ggagctgaat gccagggagg cgaagccctc gcccatgag ggctgagaag gagtgtgagc 780
atttgtgtta cccagggccg aggtgtgctg aattaccgtg cacacttgat gtgaaatgag 840
gtcgtcgtct atcgtggaaa cccagcaagg gctcacggga gatttttcca ttacaaggtc 900
gtaccatgaa aatgggtttt aaccggagtg cttgcgcttc catgctctgg caggaggagg 960
agagccacag ctgcatgtta ccgctcttgc accagctcca gaggcttggg accaggctgt 1020
ctcagttcca ggggtgctgc ggctcagacc gccctcctct ctgctctctc tctctgcctc 1080
aaatcttccc tcgtttgcat ctccctgacg cgtgcctggg ccctcgtgca agctgcttga 1140
ctcctttccg gaaacccttg ggggtgtgctg gatacagggt ccactgagga ctggagggtg 1200
ctgacactgt ggttgacccc aggggtccagc tggcgtgctt ggggctcctc tggggcatga 1260
tgaggtcaga ggggttttcc caggtgaaaa ctctgggaaa actccccagg ccatgtgacc 1320
tgccacctgc tctcccataa ttcagctcag tctgtctctc attccccac cagggtctct 1380
agctccgagg agctccccta gagggcctgg gctcaggcca gggcggtgta gtttccccac 1440
ccatgtgggg acccttgggt agtcgttga ttgggtagcc ctgaggaggc cgagatgcga 1500
tggggccagg gccgtttcca aacacagagt caggcacgtg gaaggccagg gaatccccct 1560
cccccgaggc aggagtggga gaacggagag ctgggccccg atttcacggc agccaggctg 1620
cagtgggcga ggctgtggtg gtccacgtgg cgtggggggc ggggtctgat tcaaaccgac 1680
tggggctcgg ccttctctgg ccgtgctggc cgcgcctcca cagggtctg ggggtgacgc 1740
cccgaactct agcaggtggc tatttctccc ttggaaagag agccccctac ccatgctagg 1800
tgtttccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggccttatt 1860
atttgtttaa aaacattctg ggcctggctt ccgttgttgc taaatgggga aaagacatcc 1920
cacctcagca gaggttactg gaggtgaaa ccgggggtgt ggcttgactg gtgtgatctc 1980
aggtcattcc agaagtggct caggaagtca gtgagaccag gtacatgggg ggctcaggga 2040
gtgggtgaga tgaggtacac ggggggctca ggcagtgggt gaggccaggt acatgggggg 2100
ctcaggcact ggggtgagat aggtacacgg ggggtcagg cagaggggtc gaccagggtac 2160
acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt ttcatggtag 2220
ccaggctctg gcacacctgc cccaaagtcc caggaaagct agaggccaaa gatggagggt 2280
gacagggctg gcgctggggc tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340
aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400
tatgaaaaat aaaaaaataa attagctgaa catgggtggg tgcgcctgta gttccaatac 2460
ttgggagggt gaagtgggag gatcacttga gccagggagg tggaaagtgc agtgagctga 2520
gattgcacca ctgtactgca gcctgggtga cagagtgaga gccatctca acaacaacaa 2580
agaagactga caaatgcagt ttcttggaag gaaacattta gtaggaaactt aacctacaca 2640
cagaagccaa gtcgggtgtc cgggtgtcag gagatgagat gatgggtcct cacaccatca 2700
ccccagacct aggggtttatg caccacaggg gcgggtggct cagaagggat gcgcaggagc 2760
ttgatatacy atgacatcaa ggttgtctga cgaagggcag gattcatgat aagtaacctg 2820
tggtagacaa ggaacaatgg ataaactgga aaccttagag gccttcccg aacaggggct 2880
aatcagaagc cagcatgggg ggctggcatc caggatggag ctgcttcagc ctccacatgc 2940
gtgttcatac agatgggtgca cagaaacgca gtgtacctgt gcacacacag acacgcagct 3000
actgcacac acaagcacac acacagacat gcatgcatgc atccgtgtgt gtgcacctgt 3060
gcccagagg aaacccatgc atgtgcatc atgcacgcac acaggcaccg gtggggccat 3120
gcccacaccc acgagcaccg tctgattagg aggcctttcc tctgacgtg tccgccatcc 3180
tctcag
  
```

<210> 18  
 <211> 781  
 <212> DNA  
 <213> Homo sapiens

<400> 18

```

gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctgggtg tagtgtgtca 60
ggagactgag tgaatctggg cttaggaagt tcttaccctt tttcgcatac ggaagtgtgt 120
taaccaaac actgtcaggc tcgtctgccc gccctctcgt ggggtgagca gagcacctga 180
tggaaaggac aggagctgtc tgggagctgc catccttccc accttgctct gcctggggaa 240
gcgctggggg gcctggtctc tctgttttgc cccatgggtg gatttggggg gcctggcctc 300
tctgttttgc cctgtggtgg gattggggtg tctcccgctc atggcactta gggcccttgt 360
gcaaacccag gccaaagggt taggaggagg ccaggccca gctaccccac cctctcagg 420
agcagaggcc gcgtatcacc acgacagagc cccgcgccgt cctctgcttc ccagtcaccg 480
tctctgccc ctggacactt tgtccagcat caggagggtt tctgatccgt ctgaaattca 540
agccatgtcg aacctgcggg cctgagctta acagcttcta tttctgtgtt 600
  
```

17 / 18

gtggaaattt cacctggaga agccgaagaa aacatttctg tcgtgactcc tgcgggtgctt 660  
gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccg 720  
gtgtctcctg ggaggggagc tgggctgggc ctgtgactcc tcagcctctg ttttccccc 780  
g 781

5  
<210> 19  
<211> 536  
<212> DNA  
<213> Homo sapiens

10  
<400> 19  
gcaagtgtgg gtggaggcca gtgcgggccc cacctgccca ggggtcatcc ttgaacgccc 60  
tgtgtggggc gagcagcctc agatgctgct gaagtgcaga cggcccgagg cctgaccctg 120  
ggggcctgga gccacgctgg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180  
15 agcctggcag ggtccccaac ttcttgaacc cctgtctccc atctcagggg cgatggctcc 240  
ccacgcttgg gagccttctg acccctgacc tgtgtcctct cacagcctct tccctggctg 300  
ctgccctgag ctccctgggg cctgagcaag ttctctcccc gcccgcgcgc tccagcgtca 360  
ctgggctgcc tgtctgctcg ccccggtgga ggggtgtctg tcccttcaact gaggttccca 420  
ccagccaggg ccacgaggtg caggccctgc ctgcccggcc acccacacgt cctaggaggg 480  
20 ttggaggatg ccacctctgg cctcttctgg aacggagtct gattttggcc ccgcag 536

<210> 20  
<211> 3179  
<212> DNA  
25 <213> Homo sapiens

<400> 20  
atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60  
ctgtgagtga acggggtggt ggtcagtgctg ggcctatggc ctggctgtgc atttacggaa 120  
30 gtctatgagt gaatggggtt gtggtcagtg cgggcccagc gcctggctgg gcctgggagg 180  
ttctctgagc tgtgaggcag gaggggaagg agggtagggg atagacagtg ggagccccc 240  
ccctggaaga cataacagta agtccaggcc cgaagggcag cagggatgct gggggcccag 300  
cttgggcggc ggggatgatg gagggcctgg ccagggtggc agggatgatg gggggcccag 360  
ctggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420  
35 ggtcgggccc cctcctcccc tgccctccac ctgcagcctg ggatccggat gtgttccct 480  
ggtgcacatc ctctgggcca tcagctttca tggagggtgg gggcaggggc atgacacat 540  
cctgtataaa atccaggatt cctcctctg aacgcccaca ctcaggttga aagtcacatt 600  
ccgctctgg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660  
gtgggagcgt ggagggtgtg gacacaggag gcttcagggt ggggctgtgt atgtctctc 720  
40 atctctctat catctcccag tctcatctct catctctta tcatctccca gtctcatctg 780  
tcttctctt atctcccag ctcatctgtc atctcttac catctcccag tctcatctct 840  
tactctctta tctcttagtc tcatccagac ttacctcca gggcgggtgc caggctcgca 900  
gtggagctgg acatacgtcc ttctcaggc agaaggaaact ggaaggattg cagagaacag 960  
gagggggcgc tcagagggac gcagctctgg ggtgaagaaa cagcccctcc tcagaagtgc 1020  
45 gcttgggcca cagaaaacg agggccctgc gtgagtggtc ccagagcctt ccagcagggtc 1080  
cctggtgggg ccttatggta tggccgggtc ctactgagtg cactctggac agggcttctg 1140  
gtttgagtgc agcccgagc tgctgtgtgt cggggtgggg gcttatggcc actggatatg 1200  
gcgtcattta ttgctgctgc ttcagagaat gtctgagtga ccgagcctaa tgtgtatggt 1260  
gggcccgaag ccacagactg tgcgtaaat gcactctggt gcctggagcc cccgtatagg 1320  
50 agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaac 1380  
ggaaggaggc gggcccgggc gccgtgggag gacgacctca agtgagaggt tggacagaa 1440  
agggcgggga ctctccagga gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500  
gaccaacagg tcaggccatt gttcagctat ccatcttcta caaagctcca gattcctgtt 1560  
tctccgggtg ttttttgggt aaattttact caggattact tatattttt gctaaagtat 1620  
55 tagaccctta aaaaaggat ttgctttgat atggcttaac tactaagca cctactttat 1680  
ttgtctgtt ttatttatta ttattattat tattagagat ggtgtctact ctgtcaccca 1740  
ggttgttagt gcagtggcac agtcaggtc cgtgttagcc gcaaaccccc aggtcctaag 1800  
gatcctccgg cctcagcttc ccagagtgtc gggattacag gtgtgagcca ctgcccttgc 1860  
ctggcacttt taaaaaccac tatgtaaggc cagggtccagt ggcttccaca cctgtcatcc 1920  
60 cagtagtttg ggaagccgag gcagaaggat tgcctgagc caggagtttg agaccagcat 1980  
gggtaacata gggagacccc atctctacaa aaatgcaaa aagttatccg ggcgtgggg 2040  
ccagcatctg tagtcccagc tgctcgggag gctgagtggt aggatcgctt gagcccgga 2100  
ggtcatggct gcagtggact gtgattgtac catcgactc cagcctgggc aacagagtga 2160  
gacctgtct caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaaggaa 2220  
65 gaaggaaaga gaagaagaag gaagaaggaa gaaagaagga gaaggaggcc tgctaggtgc 2280

18 / 18

taggtagact gtcaaatctc agagcaaat gaaaataaca aagtttttaa gggaaagaaa 2340  
aaccaccagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttccaca 2400  
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcagggtt 2460  
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggttgttt tcctgcctca 2520  
5 gccccacgct cctgccggtc ctgcacctgc tgtaaccgtc gatgttggtg ccagggtgccc 2580  
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640  
gcacttgtgg caggcacaat tacagccctt ccccaaagat gcccacgtcc ttctcctgga 2700  
acctgtgaat gtgtcacccg caaggcagag gctggtgaag gctgcaggtg gaatcacggc 2760  
tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggccacaa 2820  
10 gggtccttag aagtgaagga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880  
cactggccac tgctggcttt gagatggagg aggggggtccc cagccaagga atgggggcag 2940  
ccgctccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggccctgccc 3000  
acgcctcgat ttcaggccag tgggacctgt ttcagctttc cggcctccag agctgtaaga 3060  
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120  
15 cagtacaggg aaatgaatac agggacagtt ctacagtgta ctctcagccc acccctggg 3179